92d Congress }
2d Session } COMMITTEE PRINT

ENVIRONMENT, INTELLIGENCE, AND SCHOLASTIC ACHIEVEMENT

A COMPILATION OF TESTIMONY

TO THE

SELECT COMMITTEE ON EQUAL EDUCATIONAL OPPORTUNITY UNITED STATES SENATE

JUNE 1972

Printed for the use of the Select Committee on
Equal Educational Opportunity

U.S. GOVERNMENT PRINTING OFFICE
77-942 O WASHINGTON : 1972

SELECT COMMITTEE ON EQUAL EDUCATIONAL OPPORTUNITY

WALTER F. MONDALE, Minnesota, *Chairman*

JOHN L. McCLELLAN, Arkansas	ROMAN L. HRUSKA, Nebraska
WARREN G. MAGNUSON, Washington	JACOB K. JAVITS, New York
JENNINGS RANDOLPH, West Virginia	PETER H. DOMINICK, Colorado
DANIEL K. INOUYE, Hawaii	EDWARD W. BROOKE, Massachusetts
BIRCH BAYH, Indiana	MARK O. HATFIELD, Oregon
WILLIAM B. SPONG, Jr., Virginia	MARLOW W. COOK, Kentucky
SAM J. ERVIN, Jr., North Carolina	
ADLAI E. STEVENSON III, Illinois	

WILLIAM C. SMITH, *Staff Director and General Counsel*

CONTENTS

FOREWORD

The Select Committee on Equal Educational Opportunity scheduled a hearing on Environment, Intelligence and Scholastic Achievement for February 24, 1972. As that hearing could not be held as planned, statements of witnesses who were scheduled to testify are presented here along with research papers and articles on the same topic for the use of the members of the committee and others who may find this compilation of value.

WALTER F. MONDALE,
Chairman, Select Committee on Equal Educational Opportunity.

Statements and Material Submitted by Witnesses

STATEMENT OF L. I. CAVALLI-SFORZA

Mr. Chairman, and members of the committee, my name is Luigi Luca Cavalli-Sforza. I am professor of Genetics at the Stanford University Medical Center, Stanford, California. I hold an M.A. from Cambridge University (U.K.), and an M.D. from the University of Pavia (Italy). I have been Professor of Genetics at the University of Parma (Italy) from 1950 to 1960, and from that date until 1970 I was chairman of the Genetics Department at the University of Pavia. I was Vice Director, and in 1964, Director of the International Laboratory of Genetics and Biophysics at Naples. I was a fellow of the Rockefeller Foundation in 1954, and a Fellow of the Center for Advanced Study in the Behavioral Sciences at Stanford in 1970–1971. I am a foreign member of the American Academy of Arts and Sciences, a member of the American Association of Human Genetics, of the Genetical Society of Great Britain, a founding member of the Associazione Genetica Italiana, a member of the Biometric Society (of which I was President in 1966–67), and a Fellow of the Royal Statistical Society. I have been invited to give lectures and papers on human population genetics at the 10th, 11th, and 12th International Congresses of Genetics (Montreal, 1958; The Hague, 1963; and Tokyo, 1968), and at the 3rd and 4th International Congresses of Human Genetics (Chicago, 1966; Paris, 1971). I was Vice President of the International Genetics Congress, Tokyo (1968), and I am a member of the Advisory Board of the World Health Organization.

In collaboration with Prof. Walter Bodmer (then at Stanford and now Professor of Genetics at the University of Oxford, U.K.), I have written a book, *The Genetics of Human Populations*, published by Freeman & Co., San Francisco, 1971. Part of a chapter of this book was rewritten as an article, Intelligence and Race, published in *Scientific American* in October 1970, which I offer as part of my testimony. I would also like to cite an article of mine which appeared in *Disadvantaged Child* on the same subject (Problems and prospects of genetic analysis of intelligence at the intra and interracial level; Volume *3*, pp. 111–123; Ed., J. Hellmuth, Brunner Mazel, Inc., New York, 1970). These papers were published in response to Prof. A. R. Jensen's implication that the difference in IQ between ethnic groups in the U.S. is primarily genetic. I gave reasons why this assertion is unwarranted on the basis of present evidence. Little scientifically useful information has been added since those articles were written.

IQ is a measurement of certain behavioral attributes which have been chosen to represent intelligence, and which are highly correlated

with academic achievement in existing school systems. Behavior develops as a response to the environment, to which one is exposed during the whole life, of innate neurological capacities which are very poorly known or fully unknown. The relevant environments are also extremely complex and poorly known; in the case of U.S. Blacks, moreover, they are marked by well-known historical events which could only be conducive to poor socio-economic conditions, low motivation, and low self-confidence. The situation has started to change in the last years, especially since the U.S. Supreme Court decision of 1954. If the experience with the acculturation of European immigrants to the U.S. can be used as a guide, it may take up to three generations for the full benefit of these social changes to be reflected in academic performance. In fact, the environmental disadvantage of Blacks with respect to Whites has decreased only to a limited degree. On this basis, superficial analysis of the situation by the layman has created racist feelings which are not uncommon and do not contribute to ease the situation. It is also to cope with this background that social measures have to be considered.

Intelligence testing is recognized by practically all psychologists to be culturally dependent. It is based on an analysis of certain intellectual abilities developed by the individual on the basis of present educational systems and can thus at most asess the efficacy of the actual systems for developing those abilities. In fact, it correlates highly with academic success, and less highly, with economic success and related measures. It does not by itself help avoid the development of antisocial behavior, as shown by the rise in criminality, an equally important problem in modern society.

Thus IQ is an indicator of a relatively narrow side of the effects of the schooling systems, and in general of education. It is claimed by some psychologists that it is a measure of *innate* learning abilities, and this statement is based on the measurement of "heritability" of IQ. Heritability is a quantity which can be estimated in many different ways, giving numerical results which can vary widely in a given population (e.g. from 40–80% for IQ). It can also be expected to vary considerably from population to population and from time to time for the same population. The measure chosen by Jensen is the one giving the highest value and which refers exclusively to heritability *within* specific groups: English and North American Whites. It cannot give information on whether the differences *between* ethnic or even social groups is genetic or not. Moreover, examples are available that show that traits with heritabilities higher than IQ can change radically and quickly under strictly environmental effects.

The statement that ethnic or social differences in IQ are genetic means that they are to be found in the DNA. The direct analysis of DNA for these purposes is today impossible. It may become feasible in ten or twenty years from now and then the debate on this scientific issue will be less futile than it is today. We know, however, that a presumably random sample of genes, analyzed in ways that guarantees the result to be a much more close indicator of DNA diffierences than any measurable behavioral attitude can do, shows that the genetic differences between human races are, on average, rather small. They are compatible with a relatively short time of separation, during which

only a limited amount of genetic differentiation can take place. The predonderant genetic differences that are known to exist among ethnic groups (of which skin color is the most prominent example) are the consequence of adaptation to local, mostly climatic, conditions. Present knowledge gives no basis on which to draw any conclusion whatsoever on the genetic component of attributed behavioral differences among races. Whether the outcome of future studies on the manifest handicap of minority groups supports a genetic or an environmental interpretation should make little difference on basic policy in a democratic society that is constitutionally committed to equalizing the opportunities of the individual irrespective of sex, creed or skin color. The recognition that individuals or groups are handicapped should lead to measures that cure the handicap. Whether the handicap is genetic or not is entirely irrelevant to the decision of curing it. Differential education as suggested by Prof. Jensen, if it is the answer—and there are no real proofs that it is either necessary or useful—should be considered with great caution. In spite of all good intentions, it may reinforce existing psychological and social isolation between ethnic groups and above all reinforce the vicious circle whereby low socioeconomic conditions breed lower IQs and vice versa.

It seems that educators have concentrated on the present schooling system and on that narrow aspect of behavior, IQ testing. They have tended to overlook the potential of the largely unexplored area during which behavior and general attitudes may be formed, which may be important also for IQ. This is the pre-school period. Also, it seems that more attention should be paid by educators toward the development of a balanced personality, of which IQ is only one (and probably limited) facet. It is certainly not the only important one for real social progress to take place. If the apocalyptic views of a monstrous caste society, stratified by IQ, as again propounded recently by Herrnstein were correct, then educators and politicians should worry that they help avoid this degeneration. Fortunately, genetic mechanisms, for what is known of them, do not sustain a hereditary aristocracy based on objective tests of merit, which is prophesied by Herrnstein. The availablity of equal opportunity for higher education (at least to those groups that have had access to it) has been the major instrument of social mobility and of independence of parental social status. It would be very sad to see this trend reversed and the social clock set back by several centuries.

[From the Scientific American, October 1970]

INTELLIGENCE AND RACE

DO THE DIFFERENCES IN IQ SCORES BETWEEN BLACKS AND WHITES HAVE A GENETIC BASIS? TWO GENETICISTS, REVIEWING THE EVIDENCE, SUGGEST THAT THE QUESTION CANNOT BE ANSWERED IN PRESENT CIRCUMSTANCES

By WALTER F. BODMER AND LUIGI LUCA CAVALLI-SFORZA

To what extent might behavioral differences between social classes and between races be genetically determined? This question is often discussed, although generally not at a scientific level. Recently attention has been focused on the average differences in intelligence, as measured by IQ, between black and white Americans by the educational psychologist Arthur R. Jensen and the physicist William Shockley. We are geneticists who are interested in the study of the interaction between heredity and environment. Our aim in this article is to review, mainly for the nongeneticist, the meaning of race and IQ and the approaches to determining the extent to which IQ is inherited. Such a review can act as a basis for the objective assessment of the evidence for a genetic component in race and class IQ differences.

We should first define what we mean by terms such as "heredity," "intelligence" and "race." Heredity refers to those characteristics of an individual that are inherited from past generations. The primary functional unit of heredity is the gene. The human genome—the complete set of genes in an individual—consists of perhaps as many as 10 million genes. Some of these genes and their expression can now be analyzed at the biochemical level. Complex behavioral traits such as intelligence, however, are most probably influenced by the combined action of many genes. The inheritance of differences known to be determined by one gene or a few genes can be reliably predicted, but the tools for dealing with the inheritance of more complex characteristics are still relatively ineffective.

What is intelligence? A rigorous, objective definition of such a complicated characteristic is not easy to give, but for the purposes of this discussion one can focus on qualities that can actually be observed and measured. One instrument of measurement for intelligence is a test such as the Stanford-Binet procedure. Such a test is devised to measure a capacity to learn or, more generally, the capacity to benefit from experience.

Intelligence tests are based on the solution of brief problems of various kinds and on the response to simple questions. The total score is standardized for a given age by comparing it with the values of a large sample of a given reference population (such as native-born American whites). The final standardized score, which is called the intelligence quotient, is usually computed so that it is given on a scale for which

the average of the reference population is 100, and for which the spread is such that about 70 percent of the individuals have IQs in the range of 85 to 115 and 5 percent have IQs either below 70 or about 130 (corresponding to a standard deviation of about 15 points).

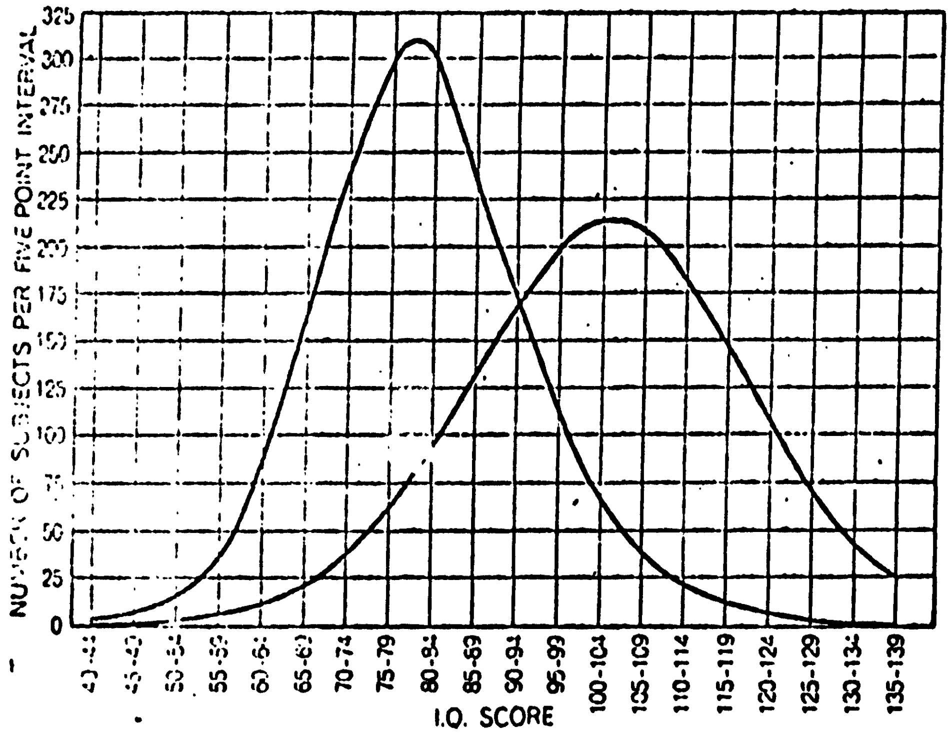
The Stanford-Binet test and other procedures yield results that correspond reasonably well to one another. More ambitious attempts have been made to measure the "general intelligence factor." There is a tendency among the more optimistic psychologists to consider such tests as measuring an "innate" or potential ability. Any given test, however, depends on the ability acquired at a given age, which is inevitably the result of the combination of innate ability and the experience of the subject. Intelligence tests are therefore at most tests of achieved ability.

This limitation is confirmed by the dependence of all intelligence tests on the particular culture of the people they are designed to test. The transfer of tests to cultures different from the one for which they were designed is usually difficult, and sometimes it is impossible. Attempts to design tests that are genuinely "culture-free" have so far failed.

A check on the usefulness of IQ measurements is provided by examining their reliability (equivalent to short-term consistency), their stability (equivalent to long-term consistency) and their validity. For the Stanford-Binet test the average difference between repeat tests after a short time interval ranges from 5.9 points at an IQ of 130 to 2.5 for IQs below 70, indicating a fairly high reproducibility. The long-term consistency of the test is less impressive, particularly if the age at the first test is lower than 5 or 6. Repetitions of testing after a period of years may show large discrepancies (up to 20 to 30 IQ points) and these differences increase with the number of years between tests.

There is at present no definition of intelligence that is precise enough to answer questions of validity in general terms. The validity of a particular test must therefore be related to its predictive aims. If the aim is to predict future school performance, then the validity is measured by how well IQ predicts that performance. The prediction will be on a probability basis, meaning that a higher IQ will usually but not always be associated with better school performance. There is, in fact, fairly general agreement that there is a high correlation between intelligence tests and success in school. The same is true for success in jobs and, in general, in society. IQ tests do have some predictive value on a probability basis, although this is limited to performance in contemporary American and European society. In this sense IQ tests do have some validity.

Races are subgroups that emerge within the same species. Like other species, the human species is made up of individuals whose genetic composition is so similar that in principle any male can mate with any female and give rise to fertile progeny. In the course of evolution this highly mobile species has spread over the entire surface of the earth. Even today, however, most individuals live out their lives within a small area. This pattern, together with geographic and other barriers, leads to considerable reproductive isolation of groups living in different regions.



IQ difference between U.S. blacks and whites emerges from a comparison of the IQ distribution in a representative sample of whites (*colored curve*) with the IQ distribution among 1,800 black children in the schools of Alabama, Florida, Georgia, Tennessee, and South Carolina (*black curve*). Wallace A. Kennedy of Florida State University, who surveyed the students' IQ, found that the mean IQ of this group was 80.7. The mean IQ of the white sample is 101.8, a difference of 21.1 points. The two samples overlap distinctly, but there is also a sizable difference between the two means. Other studies show a difference of 10 to 20 points, making Kennedy's result one of the most extreme reported.

Ecological factors, such as geology, climate and flora and fauna, may differ widely in the different habitats of a species. Natural selection, that is, the preferential survival and reproduction of individuals better fitted to their local environment, inevitably creates differences among these somewhat localized groups. In addition the isolation of one group from another allows differences to arise by the random sampling to which genes are subject from generation to generation; this process results in what is called random genetic drift.

Isolated subgroups of the same species therefore tend to differentiate. The process is a slow one: hundreds—more probably thousands—of generations may be necessary for biological differences to become easily noticeable. When sufficient time has elapsed for the differences to become obvious, we call the subgroups races.

In man biological differentiation is usually accompanied or preceded by cultural differentiation, which is a much faster process than biological evolution. The two kinds of differentiation inevitably interact. Cultural differences may contribute to perpetuation of the geographic barriers that lead to reproductive isolation. For example, religious differences may promote reproductive isolation. In the United States differences in skin color, which reflect biological differentiation, usually reduce the chances of marriage between groups. This effect,

however, is probably a direct psychological consequence not of the difference in skin color but of the parallel cultural divergence.

The relative contributions of biological and cultural factors to complex characteristics such as behavioral differences, including those that distinguish one race from another, are exceedingly difficult to identify. In this connection it is instructive to consider characteristics in which differences can easily be attributed to biological factors. It is clear, for example, that differences in skin color are mostly biological. There is a predominantly nongenetic factor—tanning—that operates during the life of an individual; it is a short-term physiological adaptation and is generally reversible. Apart from this adaptation, most of the differences in skin color both within and among races are genetic.

There are many differences among individuals that are totally under genetic control, that is, they are not subject to even the small physiological adaptation mentioned for skin color. These genetic differences are called genetic polymorphisms when the alternative versions of the genes determining them each occur within a population with a substantial frequency. Such genetic traits are generally detected by chemical or immunological tests, as in the case of the "blood groups." (There are three genes, A, B and O, that determine the ABO blood type.)

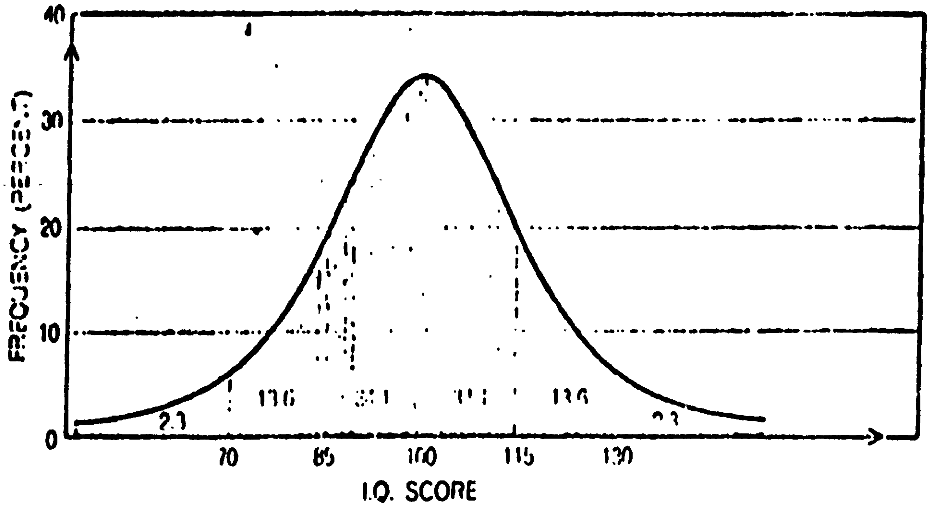
The frequencies of polymorphic genes vary widely among races. For example, in Oriental populations the frequencies of the A, B and O genes are respectively 49 percent, 18 percent and 65 percent; in Caucasian populations they are 29 percent, 4 percent and 68 percent. Such polymorphisms are a valuable aid in understanding the nature and magnitude of the biological similarities and differences among races, since they show what kinds of factor can be due solely to heredity. The inheritance of the more conspicuous face and body traits, however, is complex and not well understood, which decreases their value for the biological study of races.

The analysis of genetic polymorphisms demonstrates three very important features of the nature and extent of genetic variation within and among races. First, the extent of variation within any population generally far exceeds the average differences between populations. Second, the differences between populations and races are mostly measured by differences in the relative frequencies of a given set of genes rather than by qualitative differences as to which gene is present in any particular population. Thus any given genetic combination may be found in almost any race, but the frequency with which it is found will vary from one race to another. Third, the variation from race to race is mostly not sharp but may be almost continuous at the boundaries between races. This is the consequence of hybridization's occurring continuously at these boundaries in spite of isolation, or of the formation of hybrid groups by recent migration followed by the more complete mixing of formerly isolated groups.

As we have noted, intelligence must be a complex characteristic under the control of many genes. Extreme deviations form normal levels, as in cases of severe mental retardation, can, however, be attributed to single gene differences. Such deviations can serve to illustrate important ways in which genetic factors can affect behavior.

Consider the disease phenylketonuria. Individuals with this disease receive from both of their parents a mutated version of the gene controlling the enzyme that converts one amino acid, phenylalanine, into another, tyrosine. That gene allows phenylalanine to accumulate in the blood and in the brain, causing mental retardation. The accumulation can be checked early in life by a diet deficient in phenylalanine.

The difference between the amounts of phenylalanine in the blood of people with phenylketonuria and that in the blood of normal people, which is closely related to the primary activity of the gene causing phenylketonuria, clearly creates two gentic classes of individuals [see also illustration on this page]. When such differences are compared with differences in IQ, there is a slight overlap, but individuals afflicted with phenylketonuria can be distinguished clearly from normal individuals. This simply reflects the act that the phenylketonuric genotype, that is, the genetic constitution that leads to phenylketonuria, is associated with extreme mental retardation. If differences in head size and hair color in phenylketonuric individuals and normal individuals are compared, however, they show a considerable overlap. Although it can be said that the phenylketonuric genotype has on the average a significant effect on both head size and hair color, measurements of these characteristics cannot be used to distinguish the phenylketonuric genotype from the normal one. The reason is that the variation of head size and hair color is large compared with the average difference. Thus the genetic difference between phenylketonuric and normal individuals contributes in a major way to the variation in blood phenylalanine levels but has only a minor, although significant, effect on head size and hair color.

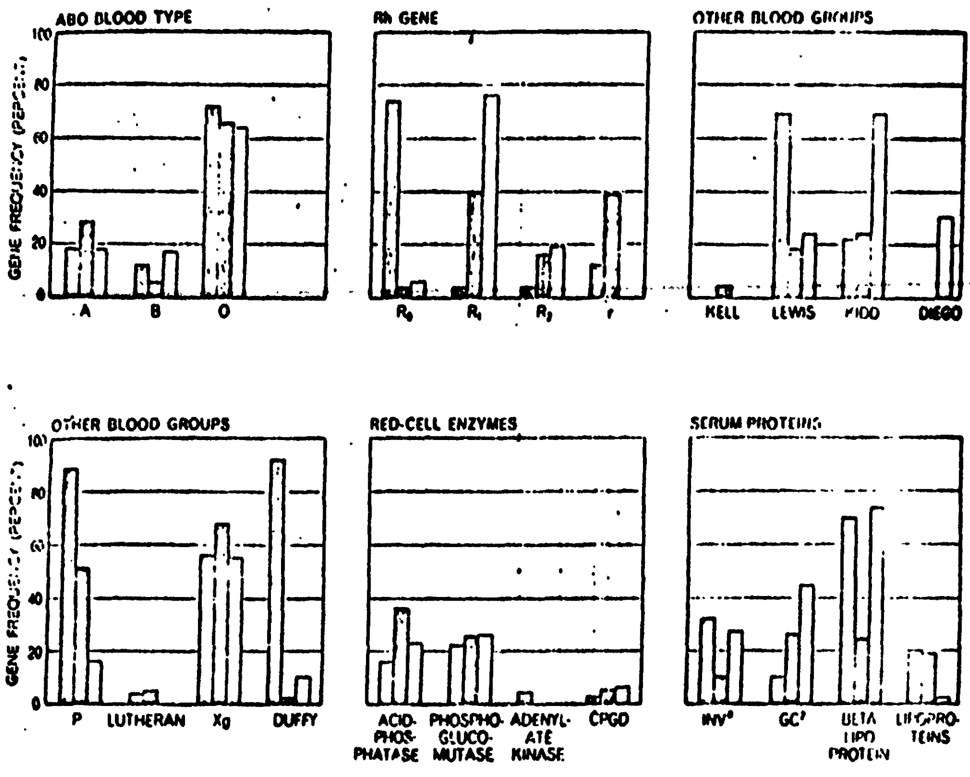


Normal distribution of IQ for a population whose mean is 100 is shown by cure. The standard deviation—that is, the usual measure of variation—is about 15 points and the distance in either direction from this mean is measured in multiples of the standard deviation. Thus about 34 percent have an IQ with a value that lies between 85 and 100, another 34 percent of the population have an IQ score of 100 to 115 points (*dark color*). Those with very high or low scores are a smaller part of population: about 2 percent have an IQ below 70, whereas another 2 percent have an IQ above 130 (*light color*).

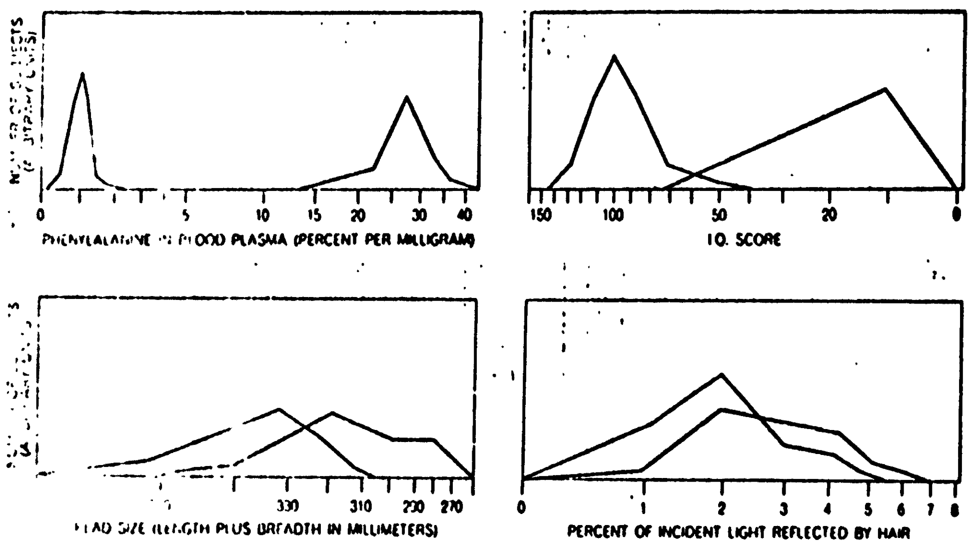
The phenylketonuric genotype is very rare, occurring with a frequency of only about one individual in 10,000. It, therefore, has little effect on the overall distribution of IQ in the population. It is now

known, however, that a large fraction of all genes are polymorphic. Among the polymorphic genes must be included many whose effect on IQ is comparable to the effect of the phenylketonuric genotype on head size or hair color. These genotype differences cannot be individually identified, but their total effect on the variation of IQ may be considerable.

The nature of phenylketonuria demonstrates another important point: The expression of a gene is profoundly influenced by environment. Phenylketonuric individuals show appreciable variation. This indicates that the genetic difference involved in phenylketonuria is by no means the only factor, or even the major factor, affecting the level of phenylalanine in the blood. It is obvious that dietary differences have a large effect, since a phenylalanine-deficient diet brings the level of this amino acid in the blood of a phenylketonuric individual almost down to normal. If an individual receives the phenylketonuric gene from only one parent, his mental development is not likely to be clinically affected. Nevertheless, he will tend to have higher than normal levels of phenylalanine in his blood. The overall variation in phenylalanine level is therefore the result of a combination of genetic factors and environmental factors. Measuring the relative contribution of genetic factors to the overall variation is thus equivalent to measuring the relative importance of genetic differences in determining this type of quantitative variation.



Frequencies of polymorphic genes among Africans, Caucasians and Orientals provide a means of differentiating these three races. (A polymorphic gene is one of a group that accounts for variability in a particular characteristic.) About half of the polymorphic systems in man are shown. Average differences in frequencies between Africans (*color*) and Caucasions (*dark color*) are 22 percent, those between Africans and Orientals (*light color*) are 30 percent, those between Caucasians and Orientals are 22 percent.



Phenylalanine levels in blood plasma shown in first set of curves at left distinguish those who carry a double dose of the defective gene that causes high phenylalanine levels (*colored curve*), a condition called phenylketonuria, from those with normal phenylalanine levels (*dark curve*). Second set of curves shows that this genotype has a direct effect on intelligence: phenylketonurics (*colored curve*) have low IQ's because accumulation of phenylalanine and its by-products in blood and nerve tissue damages the brain. Individuals with functioning gene (*dark curve*) have normal IQ's. In the third set phenylalanine levels are related to head size (displayed as the sum of head length and breadth), and in the fourth set phenylalanine levels are related to hair color (displayed as the percentage of light with a wavelength of 700 millimicrons reflected by the hair). In both cases it is obvious that the phenylketonuric genotype has a significant effect on each of these characteristics: the reflectance is greater and the head size is smaller (*colored curves*) among phenylketonurics than they are among normal individuals (*dark curves*). Yet the distribution of these characteristics is such that they cannot be used to distinguish those afflicted with phenylketonuria from those who are not.

When we turn to the analysis of a complex characteristic such as IQ, which is influenced by many genes each contributing on the average a small effect, we can expect the characteristic to be even more strongly affected by the previous history of the individual and by a host of other external, nongenetic or in any case unrelated factors, which can together be called the "environment." It is necessary to resort to statistical analysis in order to separate the effects of these various factors. Consider an experiment of nature that allows the separation, at least roughly, of environmental factors from genetic factors. This is the occurrence of two types of twins: twins that are "identical," or monozygous (derived from only a single zygote, or fertilized egg, and therefore genetically identical), and twins that are "fraternal," or dizygous (derived from two separate zygotes and therefore genetically different).

Clearly the difference between the two members of a monozygous pair is determined only by environmental factors. It would seem that the distribution of such differences among a number of pairs might tell us how much two individuals can differ because of environmental

factors alone. The members of monozygous pairs do not generally have identical IQs. The members of a given twin pair can differ by as much as 20 IQ points, although in the majority of cases they differ by less than 10. Hence environmental differences can have an effect on IQ whose average magnitude is comparable to, or slightly larger than, the difference between the IQ scores of the same individual who has been tested more than once over a period of time.

To see whether or not, and if so to what extent, genetic differences are found, we turn to dizygous twins. Here we know that in addition to the environment genetic factors also play a role in differentiating the members of a pair. The differences in IQ among dizygous pairs show a greater spread than those among monozygous pairs, indicating that the addition of genetic diversity to the purely environmental factors increases, on the average, the overall difference between members of a pair. Hence genetic factors that can contribute to the differentiation of IQs also exist among normal individuals.

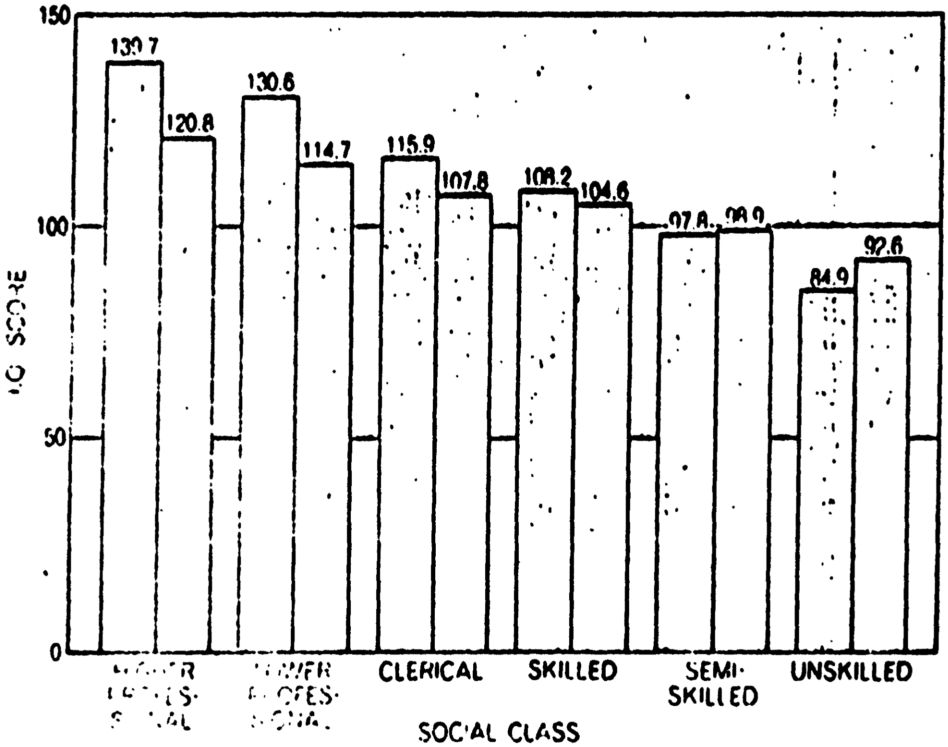
It might seem that the twin data could easily provide a measure of the relative importance of genetic variation and environmental variation. A comparison of the average difference between the members of a monozygous pair and the average difference between the members of a dizygous p... should be a good index of the comparative importance of genetic factors and environmental ones. (A minor technical point should be mentioned here. As is customary in all modern statistical analysis, it is better to consider not the mean of the differences but the mean of their squares. This is comparable to, and can easily be transformed into, a "variance," which is a well-known measure of variation.)

There are two major contrasting reasons why such a simple measure is not entirely satisfactory. First, the difference between members of a dizygous pair represents only a fraction of the genetic differences that can exist between two individuals. Dizygous twins are related to each other as two siblings are; therefore, they are more closely related than two individuals taken at random from a population. This implies a substantial reduction (roughly by a factor of two)) in the average genetic difference between dizygous twins compared with that between two randomly chosen individuals. Second, the environmental difference between members of a pair of twins encompasses only a fraction of the total environmental difference that can exist between two individuals; namely, the difference between individuals belonging to the same family. This does not take into account differences among families, which are likely to be large. Within the family the environmental differences between twins are limited. For instance, the effect of birth order is not taken into account. Differences between ordinary siblings might therefore tend to be slightly greater than those between dizygous twins. It also seems possible that the environmental differences between monozygous twins, who tend to establish special relations with each other, are not exactly comparable to those between dizygous twins. In short, whereas the contrast between monozygous and dizygous twin pairs minimizes genetic differences, it also tends to maximize environmental differences.

In order to take account of such difficulties one must try to use all available comparisons between relatives of various types and degrees, of which twin data are only a selected case. For technical reasons one often measures similarities rather than differences between two sets of values such as parent IQs and offspring IQs. Such a measure of similarity is called the correlation coefficient. It is equal to 1 when the pairs of values in the two sets are identical or, more generally, when one value is expressible as a linear function of the other. The correlation coefficient is 0 when the pairs of measurements are completely independent, and it is intermediate if there is a relation between the two sets such that one tends to increase when the other does.

The mean observed values of the correlation coefficient between parent and child IQs and between the IQs of pairs of siblings, are very nearly 0.5. This is the value one would expect on the basis of the simplest genetic model, in which the effects of any number of genes determine IQ and there are no environmental influences or complications of any kind. It seems probable, however, that the observed correlation of 0.5 is coincidental. Complicating factors such as different modes of gene action, tendencies for like to mate with like and environmental correlations among members of the same family must just happen to balance one another almost exactly to give a result that agrees with the simplest theoretical expectation. If we ignored these complications, we might conclude naively—and in contradiction to other evidence, such as the observation of twins—that biological inheritance of the simplest kind entirely determines IQ.

Instead it is necessary to seek a means of determining the relative importance of environmental factors and genetic factors even taking account of several of the complications. In theory this measurement can be made by computing the quotients known as heritability estimates. To understand what such quotients are intended to measure, consider a simplified situation. Imagine that the genotype of each individual with respect to genes affecting IQ can be identified. Individuals with the same genotype can then be grouped together. The differences among them would be the result of environmental factors, and the spread of the distribution of such differences could then be measured. Assume for the sake of simplicity that the spread of IQ due to environmental differences is the same for each genotype. If we take the IQs of all the individuals in the population, we obtain a distribution that yields the total variation of IQ. The variation within each genotype is the environmental component. The difference between the total variation and the environmental component of variation leaves a component of the total variation that may be accounted for by genetic differences. This component, when expressed as a fraction of the total variance, is one possible measure of heritability.



Social class and intelligence are closely related, a study by Sir Cyril Burt of the University of London indicates. Set of bars at left shows the mean IQ for higher professionals (*dark bar*) is 139.7, children of higher professionals have a mean IQ of 120.8 (*light bar*). Second set of bars shows that lower professionals have a mean IQ of 130.6, children of lower professionals have a mean IQ of 114.7. Third set shows that clerical workers have a mean IQ of 115.9, their children have a mean IQ of 107.8. Fourth set shows that skilled workers have a mean IQ of 108.2, their children have a mean IQ of 104.6. Fifth set shows that semiskilled workers have a means IQ of 97.8, their children have a mean IQ of 98.9. Sixth act shows that unskilled workers have a mean IQ of 84.9, their children have a mean IQ of 92.6. Mean IQ's of wives (*not shown*) correlate well with husbands. Above the mean children's IQ tends to be lower than that of parents. Below it children's IQ tends to be higher. Social mobility maintains distribution because those individuals with high IQ's tend to rise whereas those with low IQ's tend to fall.

In practice, however, the estimation of the component of the total variation that can be accounted for by genetic differences—from data on correlations between relatives—always depends on the construction of specific genetic models, and is therefore subject to the limitations of the models. One problem lies in the fact that there are a number of alternative definitions of heritability depending on the genetic model chosen, because the genetic variation may have many components that

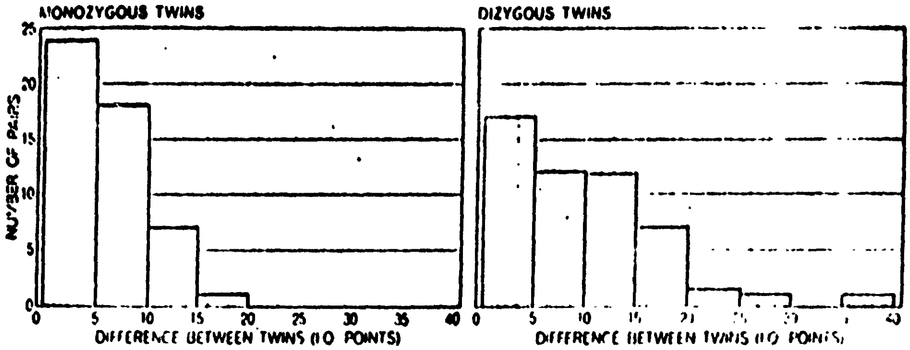
can have quite different meanings. A definition that includes only those parts of the genetic variation generally considered to be most relevant in animal and plant breeding is often used. This is called heritability in the narrow sense. If all genetic sources of variation are included, then the heritability estimate increases and is referred to as heritability in the broad sense.

The differences between these estimates of heritability can be defined quite precisely in terms of specific genetic models. The resulting estimates of heritability, however, can vary considerably. Typical heritability estimates for IQ—derived from the London population in the early 1950s, with data obtained by Sir Cyril Burt—give values of 45 to 60 percent for heritability in the narrow sense and 80 to 85 percent for heritability in the broad sense.

A further major complication for such heritability estimates has the technical name "genotype-environment interaction." The difficulty is that the realized IQ of given genotypes in different environments cannot be predicted in a simple way. A given genotype may develop better in one environment than in another, but this is not necessarily true for any other genotype. Even if it is true, the extent of the difference may not be the same. Ideally one would like to know the reaction of every genotype in every environment. Given the practically infinite variety of both environments and genotypes, this is clearly impossible.

Moreover, in man there is no way of controlling the environment. Even if all environmental influences relevant to behavioral development were known, their statistical control by appropriate measurements and subsequent statistical analysis of the data would still be extremely difficult. It should therefore be emphasized that because estimates of heritability depend on the extent of environmental and genetic variation that prevails in the population examined at the time of analysis, they are not valid for other populations or for the same population at a different time.

In animals and plants the experimental control of the environment is easier, and it is possible to explore "genotype-environment" interactions. An interesting experiment was conducted by R. Cooper and John P. Zubek of the University of Manitoba with two lines of rats in which genetic differences in the rats' capacity to find their way through a maze had been accumulated by artificial selection. The two lines of rats had been selected to be either "bright" or "dull" at finding their way through the maze. When rats from these lines were raised for one generation in a "restricted" environment that differed from the "normal" laboratory conditions, no difference between the lines could be found. Both bright and dull animals performed at the same low level. When they were raised in a stimulating environment, both did almost equally well (see illustration on page 15). Since the difference between the lines is genetic, the effect of environmental conditions should be reversible in future generations. This experiment is particularly relevant to differences in IQ because of the structure of human societies. If "ghetto" children tend to have IQ scores lower, and the children of parents of high social and economic status to have scores higher, than the level one would expect if both groups of children were reared in the same environment, then heritability estimates may be biased upward.



Experiment of nature based on IQ data collected by Horatio H. Newman of the University of Chicago gives a rough measurement of the relative influence of heredity and environment on intelligence. Chart at left shows IQ differences between the members of 50 pairs of monozygous twins, that is, twins who developed from the same egg and have identical genotypes. IQ differences between members of these pairs tend to be low: 24 pairs (or almost half of the sample) show a difference of from 0 to 5 points. Only one pair shows a difference of between 15 and 20 points. The mean difference between the members of each pair is 5.9 points. Since the genotypes in each pair are identical it appears that the environmental effect tends to be small. Second chart shows IQ differences between 15 pairs of dizygous twins, that is, twins with different genotypes who developed from separate eggs. In this case the mean difference in IQ between the members of these pairs is about 10 points. Thus a fairly large difference appears to be attributable to heredity. Such a comparison does not separate the effects of heredity and environment precisely. Members of monozygous pairs have very similar environments, whereas genotypes of dizygous twins are less different on the average by a factor of two than those of unrelated individuals. Comparison thus underestimates effect of heredity but also minimizes environmental influence.

The only potential safeguard against such bias is provided by the investigation of the same genotype or similar genotypes in different environments. In man this can be done only through the study of adopted children. A particularly interesting type of "adoption" is that in which monozygous twins are separated and reared in different families from birth or soon afterward. The outcome is in general a relatively minor average decrease in similarity. Following the same line of reasoning, the similarity between foster parents and adopted children can be measured and contrasted with that between biological parents and their children. A few such studies have been conducted. They show that the change of family environment does indeed have an effect, although it is not as great as that of biological inheritance. The correlation between foster parents and their adopted children is greater than 0, but it is undoubtedly less than that between biological parents and their offspring.

A complete analysis of such data is almost impossible because environmental variation among families and genotype-environment interactions of various kinds must be responsible for the observed effects in ways that make it difficult to disentangle their relative importance. Adoption and rearing apart take place in conditions far from those of ideal experiments, and so any conclusions are bound to be only semiquantitative. On the basis of all the available data, with allowance for these limitations, the heritability of intelligence, as measured by IQ, is still fairly high. It must be kept in mind, however, that the environmental effects in such studies are generally limited to

the differences among and within families of a fairly homogeneous section of the British or American population. They cannot be extrapolated to prediction of the effects of greater differences in environment, or of other types of difference.

There are significant differences in mean IQ among the various social classes. One of the most comprehensive and widely quoted studies of such differences and the reasons for their apparent stability over the years was published by Burt in 1961 (see illustration on previous page). His data come from schoolchildren and their parents in a typical London borough. Socioeconomic level was classified, on the basis of type of occupation, into six classes. These range from class 1, including "university teachers, those of similar standing in law, medicine, education or the church and the top people in commerce, industry or civil service," to class 6, including "unskilled laborers, casual laborers and those employed in coarse manual work." There are four main features of these data:

1. Parental mean IQ and occupational class are closely related. The mean difference between the highest and the lowest class is over 50. Although occupational class is determined mostly by the father, the relatively high correlation between the IQs of husband and wife (about 0.4) contributes to the differentiation among the classes with respect to IQ.

2. In spite of the significant variation between the parental mean IQs, the residual variation in IQ among parents within each class is still remarkably large. The mean standard deviation of the parental IQs for the different classes is 8.6, almost three-fifths of the standard deviation for the entire group. That standard deviation is about 15, and it is usual for the spread of IQs in any group.

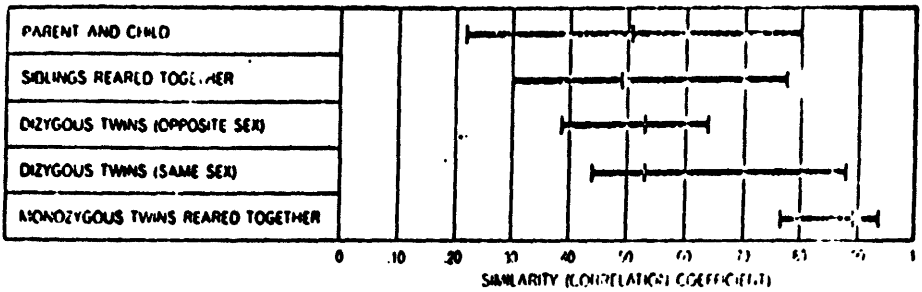
3. The mean IQ of the offspring for each class lies almost exactly between the parental mean IQs and the overall population mean IQ of 100. This is expected because it is only another way of looking at the correlation for IQ between parent and child, which as we have already seen tends to be about 0.5 in any given population.

4. The last important feature of the data is that the standard deviations of the IQ of the offspring, which average 13.2, are almost the same as the standard deviation of the general population; namely, 15. This is another indication of the existence of considerable variability of IQ within social classes. Such variability is almost as much as that in the entire population.

The most straightforward interpretation of these data is that IQ is itself a major determinant of occupational class and that it is to an appreciable extent inherited (although the data cannot be used to distinguish cultural inheritance from biological). Burt pointed out that, because of the wide distribution of IQ within each class among the offspring and the regression of the offspring to the population mean, appreciable mobility among classes is needed in each generation to maintain the class differences with respect to IQ. He estimated that to maintain a stable distribution of IQ differences among classes, at least 22 percent of the offspring would have to change class, mainly as a function of IQ in each generation. This figure is well below the

observed intergenerational social mobility in Britain, which is about 30 percent.

Fears that there may be a gradual decline in IQ because of an apparent negative correlation between IQ and fertility have been expressed ever since Francis Galton pointed out this correlation for the British ruling class in the second half of the 19th century. If there were such a persistent association, if IQ were at least in part genetically determined and if there were no counteracting environmental effects, such a decline in IQ could be expected. The fact is that no significant decline has been detected so far. The existing data, although they are admittedly limited, do not support the idea of a persistent negative correlation between IQ and overall reproductivity.



Correlation coefficients are calculations of similarity, for example, between the IQ's of two sets of relatives such as parents and children. A coefficient of 1 indicates identity, 0 indicates independence of one value from the other. These data were collected from published literature by L. Erlenmeyer-Kimling and Lissy F. Jarvik of the New York State Psychiatric Institute to derive measurements of comparative effects of heredity and environment, taking into account all possible effects of relatedness between individuals. The horizontal line at top indicates that the coefficients of samples of parents and children from different studies range from about 0.20 to 0.80. Second horizontal line indicates that coefficients for siblings reared together range from 0.30 to about 0.78. Range for dizygous (fraternal) twins of opposite sex is 0.38 to 0.65; for dizygous twins of same sex it is 0.43 to 0.88. The mean (vertical line intersecting each horizontal line) for each of these four sets of relatives is about 0.50. Monozygous (identical) twins, however, have a range of 0.77 to 0.92 with a mean of 0.89. Mean coefficient of 0.50 is that which would be expected if there were no environmental effects in IQ. Since other evidence indicates that environment exerts a significant effect, these calculations must be further refined.

The existence of culturally, and often racially, reproductively isolated subgroups within a human population almost inevitably leads to social tensions, which are the seeds of racism. This has been true throughout the history of mankind, and is by no means unique to the present tensions among different racial groups such as those between blacks and whites in the United States. Conflicts between religious groups, such as Protestants and Catholics in Northern Ireland, are examples of the same type of social tension. Cultural divergence is often accompanied by relative economic deprivation in one group or the other, which aggravates the tensions between them.

The striking outward differences between the races which is, of course the color of their skin, must be a major extra factor contribut-

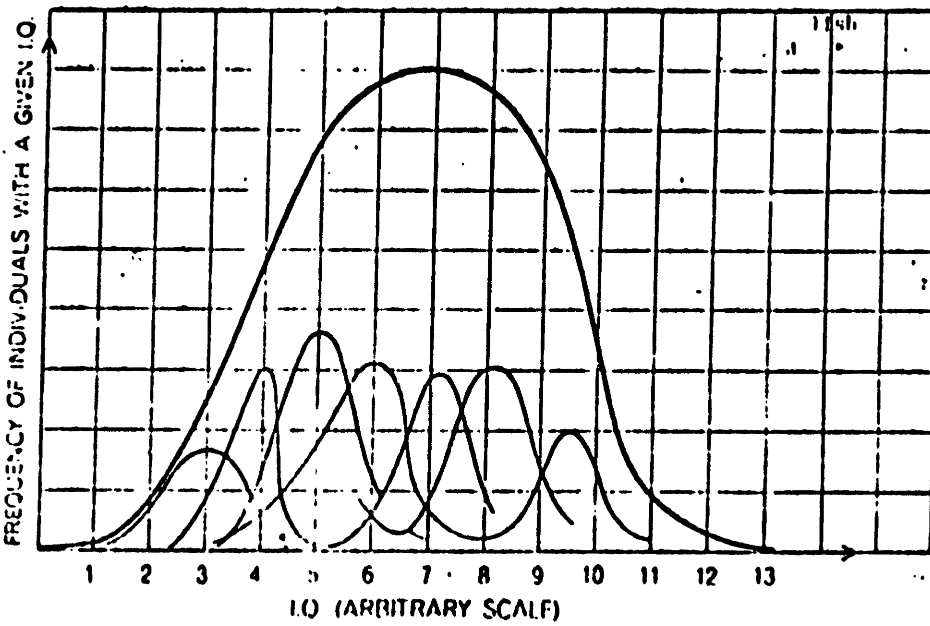
ing to the racial tensions between them. If the cultural differences between the Protestants and Catholics of Ireland disappeared, there would be no way of telling the two groups apart. The same is not true for black and white Americans. Many generations of completely random mating would be needed to even out their differences in skin color.

Such mating has not taken place in the United States. The average frequency of marriages between blacks and white throughout the United States is still only about 2 percent of the frequency that would be expected if marriages occurred at random with respect to race. This reflects the persistent high level of reproductive isolation between the races, in spite of the movement in recent years toward a strong legal stand in favor of desegregation. Hawaii is a notable exception to this separation of the races, although even there the observed frequency of mixed marriages is still only 45 to 50 percent of what would be expected if matings occurred at random.

There are two main features that clearly distinguish IQ differences among social classes described above from those between blacks and whites. First, the IQ differences among social classes relate to the environmental variation within the relatively homogeneous British population. It cannot be assumed that this range of environmental variation is comparable with the average environmental difference between black and white Americans. Second, and more important, these differences are maintained by the mobility among occupational classes that is based to a significant extent on selection for higher IQ in the higher occupational classes. There is clearly no counterpart of this mobility with respect to the differences between U.S. blacks and whites; skin color effectively bars mobility between the races.

The arguments for a substantial genetic component in the IQ difference between the races assume that existing heritability estimates for IQ can reasonably be applied to the racial difference. These estimates, however, are based on observations within the white population. We have emphasized that heritability estimates apply only to the population studied and to its particular environment. Thus the extrapolation of existing heritability estimates to the racial differences assumes that the environmental differences between the races are comparable to the environmental variation within them. Since there is no basis for making this assumption, it follows that there is no logical connection between heritabilities determined within either race and the genetic difference between them. Whether or not the variation in IQ within either race is entirely genetic or entirely environmental has no bearing on the question of the relative contribution of genetic factors and environmental factors to the differences between the races.

A major argument given by Jensen in favor of a substantial genetic component in the IQ difference is that it persists even when comparisons are made between U.S. blacks and whites of the same socioeconomic status. This status is defined in terms of schooling, occupation and income, and so it is necessarily a measure of at least a part of the environmental variation, comparable to the class differences we have discussed here.



Heritability is a measure of the relative effects of heredity and environment on characteristics such as IQ. The heritability estimate is based on the assumption that a population consists of several groups each distinguished by a different genotype and IQ distribution (*colored curves*). The total of these IQ distributions equals the IQ spread for the population (*black curve*). By definition those in each group have the same genotype, thus any variation in a group is environmental. Heredity's effect on the total IQ distribution can be calculated by averaging together the IQ spread of each group and subtracting the result from the total IQ spread.

Taken at face value—that is, on the assumption that status is truly a measure of the total environment—these data would indicate that the IQ difference is genetically determined. It is difficult to see, however, how the status of blacks and whites can be compared. The very existence of a racial stratification correlated with a relative socioeconomic deprivation makes this comparison suspect. Black schools are well known to be generally less adequate than white schools, so that equal numbers of years of schooling certainly do not mean equal educational attainment. Wide variation in the level of occupation must exist within each occupational class. Thus one would certainly expect, even for equivalent occupational classes, that the black level is on the average lower than the white. No amount of money can buy a black person's way into a privileged upper-class white community, or buy off more than 200 years of accumulated racial prejudice on the part of the whites, or reconstitute the disrupted black family, in part culturally inherited from the days of slavery. It is impossible to accept the idea that matching for status provides an adequate, or even a substantial control over the most important environmental differences between blacks and whites.

Jensen has suggested other arguments in defense of his thesis that the average IQ difference between blacks and whites is entirely genetic or mostly so, and he has challenged readers of his paper in the Har-

vard Educational Review to consider them. One is a set of data on blacks that is quite similar to those we have cited for whites, it shows the filial regression of IQ or related measurements as a function of the social class of the parents. The only conclusion one can draw is that among blacks the inheritance of IQ must also be fairly high. No conclusion can be drawn from these data concerning environmental differences between blacks and whites that affect IQ, and it is this that is the real issue.

Jensen also discusses differences between the races in rates of early motor development, and in other developmental rates, which are believed to be correlated with IQ. The argument must, by implication, be that developmental rates are determined mostly by genetic factors. Environmental influences on such rates are widely recognized, so that this information does not help to clarify the situation concerning IQ. Moreover, Jensen makes the statement, based on the well-known "Coleman report," that American Indians, in spite of poor schooling, do not show the same IQ gap as blacks. According to the Coleman report, however, American Indians typically go to schools where whites are in the majority, which is not the case for most of the schools attended by black children. (The actual difference between whites and Indians may be greater, because the sample may not have adequately represented the 70 to 80 percent of American Indians who live on reservations.) The differences between Indians and blacks or whites are clearly no easier to assess than those between blacks and whites.

Jensen states that because the gene pools of whites and blacks are known to differ and "these genetic differences are manifested in vitually every anatomical, physiological and biochemical comparison one can make between representative samples of identifiable racial groups * * * there is no reason to suppose that the brain should be exempt from this generalization." As geneticists we can state with certainty that there is no a priori reason why genes affecting IQ, which differ in the gene pools of blacks and whites, should be such that on the average whites have significantly more genes increasing IQ than blacks do. On the contrary, one should expect, assuming no tendency for high IQ genes to accumulate by selection in one race or the other, that the more polymorphic genes there are that affect IQ and that differ in frequency in blacks and whites, the less likely it is that there is an average genetic difference in IQ between the races. The same argument applies to the differences between any two racial groups.

Since natural selection is the principal agent of genetic change, is it possible that this force has produced a significant IQ difference between American blacks and whites? Using the simple theory with which plant and animal breeders predict responses to artificial selection, one can make a rough guess at the amount of selection that would have been needed to result in a difference of about 15 IQ points, such as exists between blacks and whites. The caculation is based on three assumptions: That there was no intial difference in IQ between Africans and Caucasians; that the heritability of IQ in the narrow sense is about 50 percent; and that the divergence of black Americans from

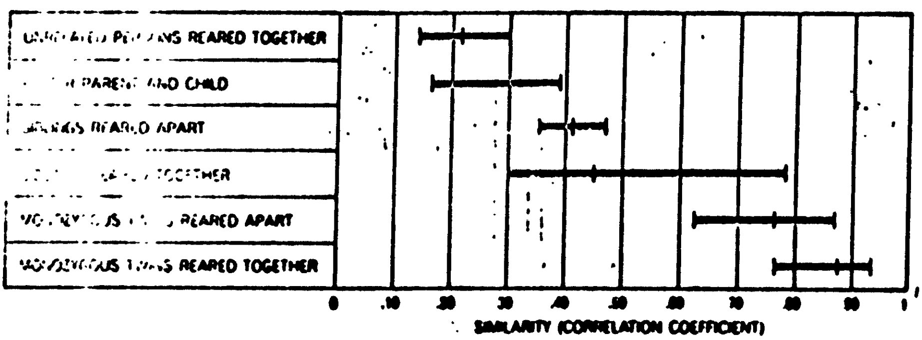
Africans started with slavery about 200 years, or seven generations, ago. This implies a mean change in IQ of about two points per generation. The predictions of the theory are that this rate of change could be achieved by the complete elimination from reproduction of about 15 percent of the most intelligent individuals in each generation. There is certainly no good basis for assuming such a level of selection against IQ during the period of slavery.

It seems to us that none of the above arguments gives any support to Jensen's conclusion. The only observation that could prove his thesis would be to compare an adequate sample of black and white children brought up in strictly comparable environments. This seems practically impossible to achieve today.

What can be said concerning environmental differences that are known or suspected to affect IQ? First it should be mentioned that, in spite of high IQ heritability estimates, the mean intrapair IQ difference found by Horatio H. Newman and his coworkers at the University of Chicago between monozygotic twins reared apart was 8 and the range was from 1 to 24. Therefore, even within the white population there is substantial environmental variation in IQ.

The following known environmental effects are also worth mentioning:

1. There is a systematic difference of as much as five IQ points between twins and nontwins, irrespective of socio-economic and other variables. This reduction in the IQ of twins could be due either to the effects of the maternal environment *in utero* or to the reduced attention parents are able to give each of two very young children born at the same time.



Effects of environment can be measured by comparing correlation coefficients of individuals with similar genetic backgrounds reared in different environments and those with different backgrounds reared in the same environment. Published data collected by Erlenmeyer-Kimling and Jarvik show that unrelated persons reared together have coefficients that range from about .15 to slightly over .30. Coefficients for foster-parents and children range from .16 to almost .40. Siblings reared apart have coefficients that range from more than .30 to more than .40. Siblings reared together have coefficients that range from .30 to almost .80. Monozygous twins reared apart have coefficients that range from more than .60 to above .80, and monozygous twins reared together have coefficients of more than .70 to more than .90. It appears that environment affects intelligence but not as strongly as heredity does.

2. It has been reported that the IQ of blacks tested by blacks was two to three points higher than when they were tested by whites.

3. Studies of the effects of protein-deficient diets administered to female rats before and during pregnancy (conducted by Stephen Zamenhof and his coworkers at the University of California School of Medicine in Los Angeles) have shown a substantial reduction in total brain DNA content of the offspring and hence presumably a reduction in the number of brain cells. The reductions were correlated with behavioral deficiencies and in man could be the basis for substantial IQ differences. There can be no doubt that in many areas the poor socioeconomic conditions of blacks are correlated with dietary deficiency. Dietary deficiencies in early childhood are likely to have similar consequences.

4. The very early home environment has long been thought to be of substantial importance for intellectual development. There are clear-cut data that demonstrate the detrimental effects of severe early sensory deprivation. There can be little doubt that both the lower socioeconomic status of U.S. blacks and a cultural inheritance dating back to slavery must on the average result in a less satisfactory home environment; this may be particularly important during the preschool years. Here again animal experiments support the importance of early experience on brain development.

5. Expectancy of failure usually leads to failure.

In his Harvard Educational Review article, Jensen chooses to minimize environmental effects such as these. We believe, however, that there is no evidence against the notion that such influences, among other environmental factors, many of which doubtless remain to be discovered, could explain essentially all the differences in IQ between blacks and whites.

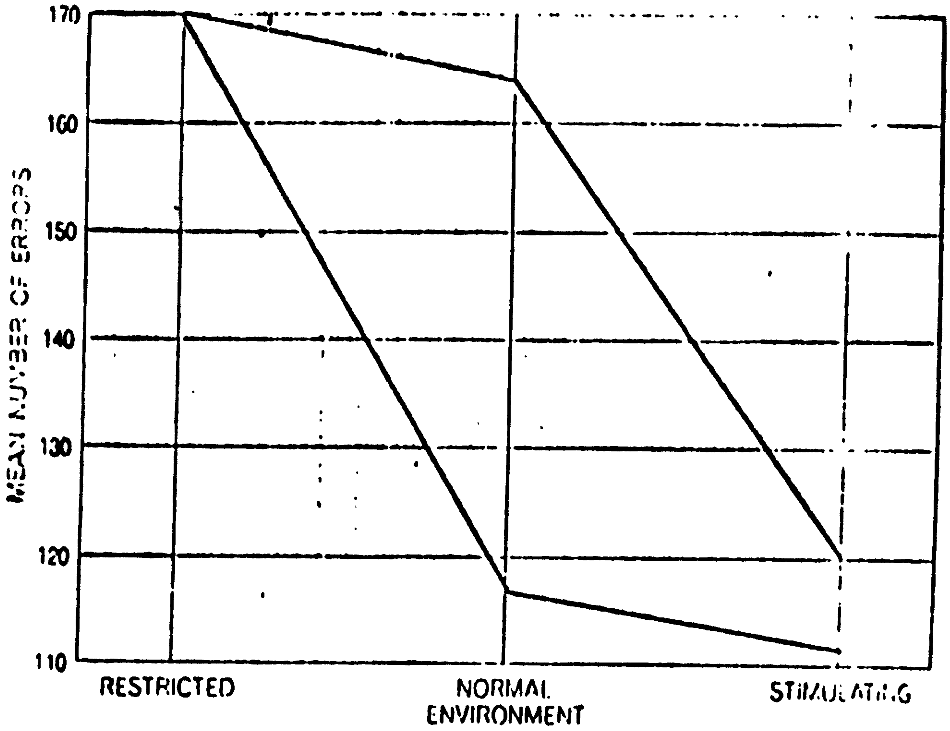
We do not by any means exclude the possibility that there could be a genetic component in the mean difference in IQ between races. We simply maintain that currently available data are inadequate to resolve this question in either direction. The only approach applicable to the study of the IQ difference between the races is that of working with black children adopted into white homes and vice versa. The adoptions would, of course, have to be at an early age to be sure of taking into account any possible effects of the early home environment. The IQs of black children adopted into white homes would also have to be compared with those of white children adopted into comparable white homes. To our knowledge no scientifically adequate studies of this nature have ever been undertaken. It is questionable whether or not such studies could be done in a reasonably controlled way at the present time. Even if they could, they would not remove the effects of prejudice directed against black people in most white communities. We therefore suggest that the question of a possible genetic basis for the race IQ difference will be almost impossible to answer satisfactorily before the environmental differences between U.S. blacks and whites have been substantially reduced.

Apart from the intrinsic difficulties in answering this question, it seems to us that there is no good case for encouraging the support of studies of this kind on either theoretical or practical grounds. From a theoretical point of view it seems unlikely that such studies would throw much light on the general problem of the genetic control of IQ, because any racial difference would be a small fraction of the total variation in IQ. The mere fact that even the relatively crude studies on the inheritance of IQ conducted so far have not taken advantage of racial differences suggests that these are not the most convenient differences to investigate. Much basic work on the biology and biochemistry of mental development under controlled conditions, making use of known genetic differences, is needed before a fuller understanding of the inheritance of IQ can be achieved.

Perhaps the only practical argument in favor of research on the race IQ difference is that, since the question that the difference is genetic has been raised, an attempt should be made to answer it. Otherwise those who now believe—we think on quite inadequate evidence—that the difference is genetic will be left to continue their campaigns for an adjustment of our educational and economic systems to take account of innate racial differences.

A demonstration that the difference is not primarily genetic could counter such campaigns. On the other hand, an answer in the opposite direction should not, in a genuinely democratic society free of race prejudice, make any difference. Our society professes to believe there should be no discrimination against an individual on the basis of race, religion or other a priori categorizations, including sex. Our accepted ethic holds that each individual should be given equal and maximum opportunity, according to his or her needs, to develop to his or her fullest potential. Surely innate differences in ability and other individual variations should be taken into account by our educational system. These differences must, however, be judged on the basis of the individual and not on the basis of race. To maintain otherwise indicates an inability to distinguish differences among individuals, from differences among populations.

We are not unaware of the dangers of either overt or implicit control of scientific inquiry. The suppression of Galileo and the success of T. D. Lysenko are two notorious examples of the evils of such control. Most investigators, however, do accept certain limitations on research on human beings, for example in the right of an individual not to be experimented on and in the confidentiality of the information collected by organizations such as the Bureau of the Census. In the present racial climate of the United States studies on racial differences in IQ, however well intentioned, could easily be misinterpreted as a form of racism and lead to an unnecessary accentuation of racial tensions. Since we believe that, for the present at least, no good case can be made for such studies on either scientific or practical grounds, we do not see any point in particularly encouraging the use of public funds for their support. There are many more useful biological problems for the scientist to attack.



Genotype-environment interaction is measured in these results from an experiment carried out by R. Cooper and John P. Zubek of the University of Manitoba. The experiment involved two strains of rats: those that were bred to be "bright"; that is, clever at finding their way through a maze, and those that were "dull". In a normal environment bright rats (*colored curve*) made only 120 errors, whereas dull rats made about 168 errors. When both strains were raised in a restricted environment, however, both made about 170 errors. When raised in a stimulating environment, both kinds of rats did equally well.

STATEMENT OF RICHARD A. GOLDSBY

Because misconception and misunderstanding of the significance of racial differences carry such great potential for harm, it is important to display and interpret the known facts about constitutional racial differences correctly and to place observations of racial differences in proper social perspective. The lessons of history illustrate the dangers that lie in an incorrect interpretation of the nature, meaning, and significance of race. Incorrect interpretation has been responsible for most tragic examples of systematic inhumanity and gratuitous intergroup violence. Just a little more than 25 years ago, offering specious arguments of Nordic racial superiority as a justification for its horrors, Nazi racism exterminated whole peoples. The incorrect and finally illegitimate doctrine of White superiority with its corollary of Black inferiority runs like an ugly thread through the entire fabric of American history. The crucial factor responsible for the introduction of such dangerous racist aberrations into the pattern of societies is the step from the recognition of palpable racial differences to an assumption of racial superiority. In the recent history of the world some racial groups have been too ready to believe that the lower technological accomplishments of other races were inherent and genetically determined rather than a reflection of life styles and the vagaries of history. This assumption of a constitutional inferiority provided the fundamental justification for the appropriation of the land of the North American Indians, the treasure of the Aztecs, and the very bodies of Blacks.

Against such a background of recorded past and possible future episodes of racially inspired malevolence it is understandable that some who are interested in bettering race relations seek to do so by denying the existence of human races. If one can prove there are no races of man, then, logically, racial bigotry and the evil resulting from its practice can be eliminated. Those who feel that the concept of race, no matter how carefully formulated, is a myth hold that supposed biological differences between races of man are, in an absolute sense, without biological reality or meaning. In popular extension, this point of view holds that apparent racial differences are superficial, only skin deep, and that men are really brothers under the skin.

From what has gone before it should be clear that this will not be the conclusion reached by this book. It is true that all men are related by their common humanity and by the common capacities that human potential confers on them, but when a six-foot Swede and a five-foot Pygmy shake hands, they are not mistaken for brothers, no matter how amicable and warm their meeting. Tall men and short men, Jews and Gentiles, for that matter men and women live in harmony not because they are alike, but because they try to respect, admire, trust,

understand, and even love each other. I do not believe it is useful or even desirable to try to promote harmony between races by denying that there are biological differences between them. Acceptance of the differences, together with an appreciation of their significance, is a more secure foundation on which to build racial harmony.

I—What Is A Race?

A race is a breeding population. A breeding population is one which for reasons of geography or culture mates largely within itself. As examples of breeding populations that we can call races we can include groups such as the Australian Aborigine and the American Black. On the other hand, it is inaccurate to speak of groups such as the Jews, who share much common culture but not a common pool of inheritance, as a race. The Australian Aborigines exist as a race primarily because of geographical factors. The isolation of their homeland, the island-continent of Australia, has for millenia insured that mating will take place with fellow islanders. The American Black is a relatively new and hybrid race maintained as a breeding population by social factors. Segregation and custom rather than geography make it likely that the majority (though the rainbowlike diversity of this group makes it clear that by no means all) of matings will be with other American Blacks. Jews, on the other hand, though a closely knit cultural group, are not a race because they are not a single but rather many different breeding populations. Two thousand years ago, in the Diaspora, the Jewish people were deliberately scattered by the Romans. In spite of the dispersion, Jews have managed to maintain their faith and have often adopted highly parallel cultural solutions to the problems they encountered as minorities. However, the racial makeup of these various Jewish populations changed, more and more diverging from each other and in most cases coming to resemble that of the people among whom they settled. The lack of racial identity of Jews is best illustrated by the diversity found in the Jewish state of Israel. There one sees a culture made up of a racial spectrum ranging from whites of northern Europe to browns from Africa and the Near East.

Members of the same race have more of their hereditary components in common with each other than with members of different breeding populations. This does not mean that all members of the same race are alike. There is enormous diversity within as well as between racial groups. It is not as though each race came from a single Adam and Eve, with every member of the race tracing his ancestry back to that pair of very great grandparents. The ancestry of all races is to some degree mixed, as though every race has had not one but multiple Adams and Eves. In most cases, too, there have been interracial matings which have from time to time introduced new and different sources of inheritance. Members of a particular race differ so much from each other because they are each the offspring of different parents each of whom traces a different lineage containing different parents butions from other breeding populations. Races are much more like stews than homogenized sauces.

Because of the diversity within races, the concept of race has most meaning when applied to populations rather than to individuals.

Imagine how little meaning it would have in terms of the over-all nature of a Mulligan stew to describe or specify in great detail the composition of a single potato or chunk of meat in the stew. Similarly one does not learn much about the nature of a race by studying a single member of the race. What one is interested in is the over-all composition—the entire recipe. We have already seen that it is possible to establish statistical criteria for recognizing racial populations. For instance, when one encounters a population in which the majority of individuals have type O, Rh positive blood, dry ear wax, and a characteristic fingerprint pattern, he can identify that population as American Indian. Paradoxically, through, if he were asked to predict the fingerprint pattern or the blood type of any particular Indian in that same population, he could not do so with certainty. If, for example, a doctor assumed, without checking, that every pint of blood from an Indian reservation was type O he would not only be incorrect but criminally negligent. It is crucial to bear in mind that racial characterizations reflect the average for a group but cannot describe the isolated individual, who is unique.

Some students of race have suggested that all of the races we see in the world today arose by combinations of a few—maybe four or five—basic racial types. This theory may or may not be true. We do not know enough about the origin of human races to say. What we can say is that there are far more than four or five major breeding populations of mankind today. Actually, even by conservative estimation, more than two dozen breeding populations and hence races can be recognized. This list includes such diverse groups as the American Indian, the Northwest European, the Southeast Asian, the Bushman, the Mediterranean peoples, the West African Black, the Australian Aborigine, and the American Black.

II—Why Discuss Race?

We do not discuss race merely to recognize and point out variations between groups. We want to see what contributions an exploration of the biology of race can make to a discussion of two larger questions, one of considerable academic interest, the other of great practical importance. First, for the satisfaction of our own curiosity we want to know why there are different races and how they might have come to be. Second, given the observed biological differences among races what, if any, is their significance for society?

In order to answer the question of racial origins it is necessary to look into the origin of man himself and consider how biological variation may have played a role in his successful adaptation to many different environments. During the course of his evolution, we know that man, like all other living things, responded to his environment by adapting biologically. Many of the characteristics such as skin color and body build that we recognize as associated with race represent adaptations evolved by man living in widely differing environments. Since his emergence in the warm, sunny climates of Africa some two million years ago man has successfully invaded and made his home in vastly different areas of the world. A key to man's global success has been the ability of different human populations to evolve varia-

tions in skin color, body structure, blood chemistry, and physiology best suited for survival in different environments.

I do not mean to suggest that every racial characteristic is or has been a significant factor in the survival of the population that it marks. In considering the origin of races, we find that a number of such "passenger" traits (nonpaying in a survival sense but along for the ride) have been included in the inheritance of racial populations. Such variations are due to two factors—the vagaries of chance and racial interbreeding. Both chance and interbreeding introduce characteristics into populations which though fascinating to those who find it amusing to classify races have no role in aiding or hindering survival.

III—The Social Implications of Race

What are the social implications of race? Exploring this question on which nearly everyone has an opinion we must recognize that there are extreme positions. At one extreme is the conviction that race is of primary and critical importance in determining the ability of an individual to function in a particular society. At the other end of the spectrum is the assertion that the notion of race is a fundamentally useless and impractical one. Recognizing that one of these positions, the one that suggests much of an individual's behavior rests on a racial foundation, is more socially dangerous than the other, both are incorrect.

As emphasized earlier, the notion of race is statistical and describes the characteristics of populations. It does not and cannot describe a particular individual in a population. Any attempt to predict the intelligence, the body build, the blood type, or even the skin color of an individual from a knowledge of his racial population is hopeless. All Northwest Europeans are not fair-skinnd, all Blacks are not black, and all Japanese are not short-statured. Given the internal variation of races, it is obvious that the general characteristics of the populations tell little about the specifics of an individual in that population.

At the same time, there are implications of practical importance that flow from the notion of race. For instance, race has important medical implications. A number of diseases are race-related. Two good examples are sickle-cell anemia and Tay-Sachs disease.

In the United States sickle-cell anemia is found almost exclusively among Blacks. This hereditary disease is characterized by a failure of the red blood cells to carry sufficient oxygen and their assumption of a sicklelike shape. The particular gene for sickle-cell anemia may be carried in a recessive or silent state by individuals who are healthy themselves. The existence of healthy carriers is accounted for by the fact that, with few exceptions, an individual carries two sets of each hereditary determinant or gene. One set comes from the father and the other from the mother. If a person carries one sickling gene but also a normal gene for the characteristic, the normal gene will be dominant. However, when two healthy carriers mate and produce offspring there is a 25 percent chance that a child produced by their union will receive a sickling gene from each parent. When an individual bears two sickling genes the debilitating and often fatal disease of sickle-cell anemia results.

Tay-Sachs disease (TSD) is a hereditary disease that results in the progressive deterioration of the brain. Afflicted individuals usually die

within two years of birth. It is found in populations of Jews who happen to be of the same race because they trace their ancestry back to Silesia and Poland. Among such Jews about one person in 40 is a healthy carrier. Again, as in the case of sickle-cell anemia, when these carriers mate there is a 25 percent chance of producing a child with TSD. Unfortunately, though research is currently in progress, there is as yet no test that will reveal healthy carriers of the TSD gene.

In addition to their medical importance, constitutional differences between racial populations have implications for mass production of such things as clothing. A garment manufacturer decides how long to make the sleeves and how to proportion the seat or the trouser length of a ready-made suit on the basis of averages. If these averages are compiled for one population and the garments are sold to another, the volume of requests for alterations is likely to be high. Many Blacks will find that the sleeves of garments modeled on White populations are too short. Oriental populations with their longer bodies and relatively shorter limbs will find these same sleeves too long. The experience of tailors performing alterations and minority populations buying ready-made clothes will confirm the fact of average racial differences in body build.

So it is true that certain features in the medical histories of populations and visits to the tailor can be race-related. What about behavior? Do some races have higher frequencies of genes for certain behavioral patterns than others, just as some populations have higher frequencies of genes for sickle-cell anemia or Tay-Sachs disease? The answer depends on the type of behavior. It is a type of behavior which depends on traits whose gene frequencies vary from one racial population to another, then the answer is yes, definitely. Average population height is a good example of a race-related genetically determined trait that is important in the type of behavior known as basketball playing, an activity that places an extraordinary premium on height. Consequently taller populations such as certain Caucasoid and Black groups have a distant genetic advantage for the practice of this behavioral pattern over shorter groups such as Mongoloids or Bushmen. And while a particular Japanese may be a superior basketball player, we expect taller populations to produce more Jerry Wests and Wilt Chamberlains.

On the other hand the decathlon is a tournament that samples an exceedingly broad spectrum of abilities including speed, endurance, body poise, and even the personality trait of determination. In an event which tests such a wide range of different abilities the particular inborn strengths of a population for one event tend to be balanced off by its inborn weaknesses for another. Hence, a decathlon winner is a rare combination of balanced excellence that is as likely to be produced by one race as another. We are not surprised to find that the list of decathlon stars includes Bob Mathias (a Caucasoid), C. K. Yang (an Oriental), and Rafer Johnson (a Black).

There are, of course, many types of behavior which do not lend themselves to such simple analysis of race-relatedness. There are behavioral characteristics which, though substantially inherited, depend upon a suitable environment for their fullest development. When one attempts to make interracial comparisons of such environment-sensitive traits between racial groups whose experience and environment differ, great

difficulty is encountered in deciding to what degree the observed difference is caused by nature and therefore genetically determined and how much is a result of nurture and hence a reflection of culture and experience.

Performance on IQ test is an example of genetically determined behavior which is environment-sensitive. Interracial differences in performance on IQ tests are matters of documented fact. Here in the United States, for instance, Black populations score lower on IQ tests than White populations. Does this mean that Blacks are of inherently lower intellectual potential than Whites or does it reflect a difference in the experience or motivation of these groups? Guesses, some of them informed, have been made but no one really knows what the answer would be if Whites and Blacks shared essentially common backgrounds. The fact is that at present these groups by and large live apart, attend different schools, are motivated by different factors, and honor different values. Until all of these factors can be properly evaluated, an explanation for the observed differences in IQ scores must be delayed.

These considerations of the relationship between race and behavior point out that studies of human variation cannot produce conclusive answers when critical bits of information are unknown or unavailable and thus serve to remind us of an important point. In considering the subject of race and races understanding depends not only on knowing what is known but also in realizing what is not.

STATEMENT OF IRVING I. GOTTESMAN

Mr. Chairman and members of the committee, my name is Irving I. Gottesman and I am a Professor of Psychology at the University of Minnesota and a Director of the Behavior Genetics Training Program there. Although I am Chairman of the Public and Professional Affairs Committee of the newly-formed Behavior Genetics Society, I do *not* appear in that capacity today. Born to Jewish-Hungarian immigrants, I was reared in the Hough-Glenville district of Cleveland, Ohio and attended a "quality integrated" elementary school (Miles Standish) in the 1930's without knowing until much later that it deserved a special label. I participated fully in the American Dream, graduating from Shaker Heights High School, earning a B.S. from Illinois Institute of Technology, serving as a naval officer during the Korean War, and then becoming a Ph.D. at the University of Minnesota. I have had the privilege of teaching at Harvard University and the University of North Carolina and was a USPHS Special Fellow in psychiatric genetics at the Medical Research Council Unit in London (1963–64). I have been concerned with the role of genetics in human adaptive behavior, personality, and mental disorders for the past fourteen years; in passing, I have personally studied more than 300 pairs of twins. My publications in these areas include three books and numerous scientific articles including a chapter, "Biogenetics of Race and Class," in the book edited by Deutsch, Katz, and Jensen, *Social Class, Race, and Psychological Development* (1968). I am an Associate Editor of the *American Journal of Human Genetics*, *Social Biology*, and *Behavior Genetics*, and for five years served in similar capacity for *Monographs of the Society for Research in Child Development*.

Closed minds, of whatever persuasion, are unwarranted on the nature-nurture aspects of human behavior, and serve mainly to aid and abet those who must rationalize their prejudices. It is unfortunate but true that many of the facts generated by our research on behavior can be misused by racists and bigots in support of *apartheid* mentality. Knowledge, like fire, can be used for good or evil, but few people wish that fire had never been invented. The political distortion of sound biological ideas in the past has given the science of human genetics a bad reputation which it is trying to overcome.

Professor Marvin Bressler, the noted Princeton sociologist, has observed that "The capacity of any minority to accept serenely the fact of resemblances and differences among people presupposes that the majority will not frustrate their aspirations for full emancipation. There are differences in average IQ scores among white ethnic groups, but the origin and nature of Irish characteristics, for example, are not a matter of public debate or anxious self-scrutiny, precisely because

a once-persecuted minority has now been absorbed—despite an unfavorable stereotype—into the mainstream of American society. In this connection, [some] seem to believe that any appearance of scientific support of a doctrine of intrinsic race differences might have unfortunate effects on the attitudes of the white population. It would presumably reduce the fervor of the virtuous, justify the passivity of the uncommitted, and provide moral succor to the bigot. These defensible expectations must be balanced against countervailing considerations. An ideology that tacitly appeals to biological equality as a condition for human emancipation corrupts the idea of freedom. Moreover, it encourages decent men to tremble at the prospect of "inconvenient" findings that may emerge in future scientific research. This unseemly anti-intellectualism is doubly degrading because it is probably unnecessary" (Bressler, 1967).

In preparing for my appearance here today I struggled with the problem of defining "equal educational opportunity"; I believe the definition has bedeviled and eluded you, for good reasons. The Declaration of Independence was too succinct when it said "that all men are created equal." Equal in that context meant deserving equal protection of the law and the inalienable right to the equal opportunity for developing their capacities to the fullest practical extent. It could not have meant that the capacities of men are identical; the distinction between equality and identity helps, I believe, add to perspective. The genetic diversity which characterizes our species is a necessary and marvelous part of our evolutionary heritage which has permitted us to survive.

If we shift our attention to the concept of *equal nutritional opportunity*, for a minute, it will highlight some of the problems with equal educational opportunity. The goal in this instance being to guarantee freedom from starvation as a minimum and *optimum* nutrition as an ideal. You can now see that *equality of inputs* will not guarantee our goals for nutrition (or education); some individuals would get too much food and some not enough. You can also see that *equality of outputs* is an inadequate and unreasonable criterion for the presence of equal nutritional opportunity since we cannot mandate identical height and weight for men, women, and children, just as we cannot mandate identical achievement levels at the end of compulsory education ages. *Equal availability of resources* so that an individual *could* optimize his nutritional needs is the one sense of equal nutritional opportunity that permits a pragmatic definition; and so too with education, I believe. Unfortunately there is nothing comparable yet to a sex, height, age, and body build chart in the field of education which would permit prescribing optimums for education, but efforts are being made. I like the nutrition analogy to education for other reasons too. It implies that different menus can be equally effective, that variety is necessary to avoid boredom, that some individuals may be allergic or unable to cope with certain subjects but tolerate others, and that some will show special aptitudes for some subjects and not others.

The great humanitarian and evolutionist Dobzhansky has summarized these issues as well as anyone: "It is a perversion of the ethic of equality to endeavor to reduce everybody to a uniform level of achievement." "From each according to his ability" is the famous motto of Marxian socialism, and it behooves democracy to grant no less recognition to the diversity of human individualities. This is not an apology

for "rugged individualism"; the "ruggedness" amounts often to indifference or even contempt for individualities of others. Equality is, however, not an end in itself but a means to an end, which can only be the self-actualization of human individuals and the fullest possible realization of their socially valuable capacities and potentialities. Individuals and groups will arrange their lives differently, in accordance with their diverse notions of what form of happiness they wish to pursue. Their contributions to mankind's store of achievements will be different in kind and in magnitude. The point is that everyone should be able to contribute to the limit of his ability. To deny the equality of opportunity to persons or groups is evil, because it results in wastage of talent, ability, and aptitude, besides being contrary to the basic ethic of humanity.

The current state of knowledge about the origins or causes of the observed racial and social class differences observed in IQ test scores, and educational and social attainment does *not* deserve any reverence. Both scientists and the general public who manage to cut through the thick fog of rhetoric surrounding these inflammatory issues are then confronted by a welter of contradictory facts, near facts, contradictory opinions about the same facts, scientific opinions, and opinions of scientists. The public is bound to be confused by the distinction between the assertion of a hypothesis and the assertion of an established truth. The host of caveats and underlying assumptions which should accompany information on human differences and similarities are seldom made explicit. A scientist may choose to behave like a lawyer in an adversary proceeding and only select and present the data which support one point of view; there is no law against this style of writing. I am conflicted about the benefits of scientists conjecturing in "public" about the alleged genetic basis of the difference between intelligence test scores of blacks and whites in this country, ever mindful of academic freedom, Galileo, and Lysenko. But, could this be an instance (*pace* Holmes) analogous to shouting "FIRE! . . . I think" in a crowded theater, as suggested by my colleague Professor Scarr.

On my desk I have a philosophical caption by the Peanuts cartoonist Schulz; it reads, "There's no heavier burden than a great potential!" In my view the social and behavioral sciences are in that predicament—they have a great potential, unrealized as yet, for contributing to the solutions to the problems surrounding Equal Educational Opportunity. Right now, however, most of the information we have about learning, cognition, motivation, genetic aspects of human behavior, forgetting, and achievement is in the category of basic science and is far from being translatable into useful applied knowledge for the problems of society. We have accumulated a lot of data about maze running in rats, reinforcement schedules in pigeons, and eyelid conditioning and nonsense syllable learning in college sophomores, but no definitive knowledge about how language, thought, or reading with all their complexities are acquired and perfected. The neurophysiology and molecular biology of learning are virtually *terra incognita*. As one of our foremost theorists Professor William Estes said in 1970, ". . . a discipline having to do with the development of learning ability has only begun to take form." I have no intention of belittling the basic accomplishments, but I do bemoan the fact, for example, that our nearest primate relative, the chimpanzee, has hardly been tapped as a

source of information. The structure of higher education in this country speaks to the split between the applied and basic aspects of human education. We have Schools of Education administratively and intellectually isolated from Graduate Schools of Psychology, even to the point of offering different doctoral degrees (Ed.D. vs. Ph.D.). Behavioral genetics is to education as Physics is to structural engineering; one doesn't ask a behavioral geneticist how to educate a child nor a physicist how to build a bridge, although each provides information which feeds in to such applied endeavors. The National Institute of Education proposed in President Nixon's speech of March 24, 1970 to conduct and sponsor basic and applied educational research has the potential for alleviating some of the woes of equal educational opportunity.

From the vantage point of a behavioral geneticist, I believe I can make a useful contribution to the deliberations of this Committee by citing some widely held opinions about race, class, and intellect and then commenting on the faults and virtues of the propositions.

1. The IQ Test Measures Innate Intellectual Ability

Not true. At best it is an indicator of it, the result of a particular genetic endowment developmentally (ontogenetically) elaborated by successive interactions with relevant pre and post natal environmental events. IQ is a phenotypic characteristic; we cannot measure the genetic capacity of an individual for behaving intelligently. However, the IQ test is a good predictor of success in school for groups of middle-class children of whatever race. IQ is not the same as achievement; in fact it only accounts for some 25% of the variance in various abilities. Children matched for IQ do not perform at the same level on achievement tests (Baughman and Dahlstrom, 1968). I like to think of IQ, when validly assessed, as analogous to the horsepower of a car. It tells you one of many useful facts, but is not sufficient to permit prediction of the performance of the car over its lifetime; a VW and a Ferrari will both do the job of getting you from New York to San Francisco. Professor Jensen has made the good point that you don't throw away a thermometer just because it tells you you have a fever and you don't want one. I agree and urge that accurate individual assessment form the basis for individualized instruction. However, there are at least two situations I can imagine where you would not take action as a result of the thermometer reading. If, unknown to the examiner, a child had been sucking on ice cubes or drinking hot tea before testing, you would be obtaining accurate but misleading information. I would suggest to you with respect to the IQ testing of many disadvantaged children, that the readings reflect an intellectual diet of ice cubes between the time of conception and entrance to elementary school. One last point about the complexity of human intellect: IQ is a useful but greatly oversimplified means to assess the 120 or so components of it posited by Guilford.

2. The 10 to 20 IQ Point Differences Between Large Samples of Black and White Children Cannot be Taken as Proof of "Genetic Inferiority"

True. Although genes could be involved to some extent, the phenotypic differences do not compel a genetic explanation. Even in the

case of identical twins when the genes in common are known to be the same, very large differences in IQ scores and achievement have been observed for some pairs whether they have been reared together or apart. Professor Jensen has made a case to the effect that the 20 or 24 point IQ differences between identical twins are compatible with the hypothesis that they could have occurred by chance. I would put it otherwise. The large differences between individuals matched genetically are associated with systematic differences in the environment, prenatally and postnatally, including educational opportunity, social class, physical factors, and personality factors. My London colleague James Shields who has personally studied more pairs of identical twins reared apart than anyone else has made similar suggestions. Whether the differences are ultimately shown to be due to chance or systematic environmental effects, the fact remains that both interpretations are compatible with the data. Compatibility between data and hypothesis is necessary but does not constitute proof; an impartial judge would need more data to reach a verdict on which of us, if either, is correct. To the extent that it is valid to invoke evolutionary theory in this debate, it would not lead us to expect a difference between groups, although separated in time and space, if the trait under consideration were under the same kind of selection pressure (i.e., conferred the same selective advantage). Parallel rather than divergent evolution would have taken place. It is difficult to conceive of any human environment in which poor adaptive behavior or unsuccessful problem solving would have given an advantage to their gene carriers.

3. Even When Blacks and Whites Are Matched for Social-Economic Status, the Blacks Obtain Much Lower IQ Scores

Not so, according to the most recent careful work in this area by Dr. Paul Nichols. He used 5,256 white and 4,613 black 7-year-old children who formed the Boston, Baltimore, and Philadelphia samples of the NINDS Collaborative Project. By city, the parents were matched for education, occupation, and income. The results for their children: in Boston, average IQ for whites and blacks, 104.2 vs. 100.0; in Baltimore/Philadelphia, for whites and blacks, 95.3 vs. 91.2. It is only when you contrast Boston whites with Baltimore/Philadelphia blacks that the typical 15-point deficit is found. Very similar data were obtained when the same children were tested at age 4.

4. Bona Fide Mental Retardation and IQ Scores Under 90 Would Remain Even With Equal Educational Opportunity

Sad, but true. Look at a country with pragmatic socialism like Sweden. Two per cent or so of the population are mentally retarded and have IQs less than 70 just like in this country. The distribution of scores is virtually identical with that for the white U.S. population with a mean of 100. However, the IQ scores of disadvantaged children is a poor predictor of bona fide mental retardation.

5. Like Begets Like, or Like Father, Like Son

Cats do have kittens rather than puppies, but the son of a ditch digger is not destined by his genes to have his father's occupation

or IQ. If we look at the children of semi-skilled workers (Burt, 1961), for example, they range in IQ all the way from 55 to 135 with an average of 98.9. The workers themselves ranged from 65 to 125 with an average of 97.8. These kinds of data reveal the fallaciousness of "typological thinking"; knowing the average for a group of individuals does not permit you to state the value for an individual in that group, whether the group be defined simplistically by class or color.

6. Talent is Rewarded by Upward Social Mobility

By and large this is correct for the U.S. white population, but we do not know to what extent it may be true for the black population. Dr. J. Waller (1971) has shown that of sons averaging 30 IQ points higher than their fathers, 67% moved to a higher social class; of sons averaging 30 IQ points lower 88% moved to a lower social class. The same phenomenon of social mobility as a function of ability has been observed in England, but to a lesser degree. A total of 55% of sons in Waller's sample changed their social class compared to their fathers in both the upward and downward directions. The net result of an open-class democratic system combined with equality of opportunity and assortative mating is a process which in the long run guarantees genetically based differences among social classes. When blacks are permitted to have social mobility as a consequence of their talent, stratification similar to that in the white population is likely to result.

7. The Degree of Genetic Determination of IQ Scores is Generally High

Yes, in the white populations in which it has been calculated it runs between 50% and 80%, but so what? The statistic called heritability characterizes a trait only under the conditions of history of that sampled population. The value can be reduced by changing from an expressive to a suppressive environment, as well as increased by changing to a more trait-relevant expressive environment. One might say, heritability today and gone tomorrow.

8. Large Changes in Trait Values Require Large Changes in Gene Pools

Not correct. In 1960, 9% of black families had incomes greater than ten thousand dollars a year; in 1970 the percent rose to 24%. In 1960, 38% of all blacks aged 25–29 had completed high school; in 1970, the percent rose to 56%. In 1900, the life expectancy of a black male was 32, in 1940, it was 52, and in recent times about 61 (still six years less than for a white). Such dramatic changes in "Achievements" in so short a time cannot reasonably be attributed to genetic changes, but they are compatible with the conclusion that the black population in this country is making increasing use of the opportunities available. It appears as if even more tolerance and more civil rights legislation will be required after the day of equal educational opportunity has arrived: the average income of a white man with a high school education is $8,631 while a black man with the same amount of education has an average income of only $6,144.

In conclusion, I hope the deliberations of this Select Committee on Equal Educational Opportunity will be blessed by luck and wisdom. The American Dream will be so much more vital and fascinating when it is in multicolor red, brown, black, and yellow, instead of monochromatic white.

BIOLOGY, SOCIAL STRUCTURE, AND EQUALITY

(Talk to National Conference on Social Welfare, Chicago, June 2, 1970.)

By Irving I. Gottesman, University of Minnesota

I am a scientist and not a politician or a social activist. If you believe that knowledge which can be misued by bigots or racists should be withheld from public knowledge, I will have reason to regret my acceptance of Therese Lansburgh's invitation to step down from my ivory tower so as to further interdisciplinary communication. As a member in good standing of both the American Civil Liberties Union and the American Eugenics Society, and as a university professor with a reputation as an honest broker between the science of human genetics and the science of human behavior, I feel that I can share with you some of my perspectives (Gottesman, 1968, 1970) on the biogeneics of race and social class with relative immunity.

Not too long ago in the the British House of Commons a question was put to the Minister for Education. A member of Parliament said to him, "Sir, it has come to my attention that 50 percent of the people in this country have IQ scores less than 100. What do you propose to do about it?" I didn't expect all of you to laugh at that humorous anectdote because the basic biological variability of many human characteristics is not generally appreciated.

Seventy years ago at the turn of the century Gregor Mendel's (1822–84) revolutionary ideas about the genetic basis for the transmission of physical resemblances between parents and offspring were rediscovered and widely circulated—some 35 years after the monk's garden experiments. The awareness of genetics coincided with the publication of the equally revolutionary ideas of Sigmund Freud's (1856–1939) "The Interpretation of Dreams." Independently of these major developments, Charles Darwin (1809–82) had revolutionized thinking about man's status in the universe with his theory of evolution enunciated in "The Origin of Species" (1859). These three movements were destined to overlap, intertwine and greatly influence thinking about the basic nature of mankind as the 20th century unfolded. Behavioral sciences, dominated by the more prestigious natural sciences eagerly embraced, applied and misapplied the neat principles of Mendelian genetics to their burdensome, unsolved "diseases"—insanity, feeblemindedness, and alcoholism. Regrettably, unwarranted eugenic fervor, social Darwinism and racism were stimulated by the political distortion of sound bological ideas; I categorically disassociate my thinking from such evils (cf. Dunn, 1962; Haller, 1963).

My first figure shows the current distribution of IQ test scores in the present population of white-skinned people. Let me state at the outset that IQ test scores cannot routinely be equated with intellectual

ability or capacity. The average or mean score is 100 and the range goes from about IQ 50 to IQ 150. The distribution is shown by the shaded lines. The other curve you see in the figure represents the theoretical distribution of IQ scores if everyone had the same genes with regard to intellectual potential and all variation in the population were only due to environmetal factors.

The next figure is of more interest even though it is still theoretical because it is relevant to prenatal and early childhood intervention. In this next figure you again see in the shaded area the current distribution of IQ scores. The superimposed curve shows you what the theoretical distribution would be if all environmental factors were equalized and each child was exposed to the "average" environment. This would mean that all the remaining variation is due to genetic factors. You can see that the effect of having everyone exposed to the average environment results in the elimination of some very low and some very high IQ scores, but it leaves the essential distribution intact. This curve should make it clear why the question to the minister of education should have resulted in a laugh.

The widely held notion that the eradiction of the dehumanizing conditions which infest our and other nations' ghettos would lead to the elimination of individuals with low IQ scores is not supportable.

The next figure shows you the empirically obtained IQ score distributions for a sample of 1,800 Southeastern Negro elementary schoolchildren and for the normative sample on the Stanford-Binet. The children in grades 1 through 6 were representative of the population of black children in Florida, Georgia, Alabama, Tennessee, and South Carolina. As you can see, the average IQ score was about 81. The average for the 1960 normative sample of white children was about IQ 102. Notice at the same time that the overlap of the two distributions is virtually complete. The major question about the differences between the two distributions is, "Do they represent differences in the genes or differences in the environments?"

Marvin Bressler, a Princeton sociologist, has noted that "The capacity of any minority to accept serenely the fact of resemblances and differences among people presupposes that the majority will not frustrate their aspirations for full emancipation. There are differences in average IQ scores among white ethnic groups, but the origin and nature of Irish characteristics, for example, are not a matter of public debate or anxious self-scrutiny, precisely because a once-persecuted minority has now been absorbed—despite an unfavorable stereotype—into the mainstream of American society. In this connection, [some] seem to believe that any appearance of scientific support of a doctrine of intrinsic race differences might have unfortunate effects on the attitudes of the white population. It would presumably reduce the fervor of the virtuous, justify the passivity of the uncommitted, and provide moral succor to the bigot. These defensible expectations must be balanced against countervailing considerations. An ideology that tacitly appeals to biological equality as a condition for human emancipation corrupts the idea of freedom. Moreover, it encourages decent men to tremble at the prospect of "inconvenient" findings that may emerge in future scientific research. This unseemly anti-intellectualism is doubly degrading because it is probably unnecessary" (Bressler, 1967).

What kinds of concrete meaning can be attached to the observed average differences of 10 IQ points between groups of Northern Negroes and whites and 20 IQ points between groups of Southeastern Negroes and whites? As a result of a kind of overselling of the practical uses of IQ tests, professionals and laymen alike appear to invest test scores with an undeserved aura of permanence and profound significance. An exposure to data on the construct validity of intelligence tests and the susceptibility to change of IQs (Maher, 1963; Hunt, 1961) would help temper this naive enthusiasm. It is too easy to forget the operations by which an IQ is computed. For example on the 1937 Binet test the answer to a question is most often worth 2 months of mental age credit; the answer to one question is thus good for 2 or 3 IQ points. Given two 8-year-old children with IQs of 90 and 100, the latter has been able to answer five more questions correctly than his classmate. It should be obvious that when an IQ test has fewer total questions than the Binet, each correct answer is worth proportionately more than 2 or 3 IQ points.

A vast clinical and experimental literature has grown up documenting the importance of early experience for later development for both animals and man (e.g., Hunt, 1961; Brackbill, 1964). This literature strongly suggests that perceptual and stimulus deprivation of a rather subtle nature is capable of handicapping subsequent development. None of these statements should yet be taken to mean that true mental deficiency can be cured by a program of enriched education. One of my goals is to explain Negro-white differences in IQ rather than to explain them away. (cf. Cravioto, 1968; Deutsch, 1969).

Another way to gain perspective about the meaning of a 10 or 20 IQ point difference is to look at the data on within pair differences in intelligence for identical (MZ) and fraternal (DZ) twins. The reason why these data are important to the issue of race differences in intelligence is because some people have interpreted the mean differences observed between white and Negro American samples as sufficient evidence of "genetic inferiority" or of differential capacity for intelligence. Since a pair of identical twins has no difference in their genes (they come from one egg which has split in two) any differences between them must be due to the environment, either prenatally or postnatally. If we construct two samples of identical genetic constitution by taking one member of each pair of identical twins in one group and one in the other, what kind of a (absolute) mean IQ difference do we find? Even though the gene pools do not differ and even though each of the two groups has been raised under more or less the same regime, the mean difference amounts to 6 IQ points for the sample of 50 pairs studied by Newman, Freeman, and Holzinger (1937). The range of within pair differences was 0 to 20 points. Thus, even when gene pools are known to be matched, appreciable differences in mean IQ can be observed that could only have been associated with chance or systematic environmental differences.

A better appreciation of the influence of the environment on IQ can be gained from looking at the two unique samples of thoroughly described identical twins who have been reared apart and thus in potentially discriminably different environments. Such data are crucial to understanding the range of intelligence which can be manifested

by persons of the same genetic background. In the 19 pairs of identical twins reared apart studied by Newman et al. (1937), the average intrapair difference on the Binet was 8 IQ points. The range of differences however was from 1 to 24 points. A very similar picture is given in a remarkably large sample of 38 pairs of identical twins reared apart and studied by Shields (1962). When the tests used in this larger study are converted into IQ point equivalents (Shields & Gottesman, 1965), the average intrapair difference for the identicals is 14 points on a verbal IQ test and 10 points on a nonverbal test. The corresponding differences for a control sample of 34 identical pairs reared together which Shields studied with the same instruments were 9 IQ points for both tests. At least 25 percent of the sample of identicals reared apart had within pair IQ point differences on at least one of the tests which exceeded 16 points.

It is obvious from looking at the data on identical twins that individuals with exactly the same genetic constitution can differ widely on the phenotypic trait we measure with IQ tests and label intelligence. The differences observed so far between whites and Negroes can hardly be accepted as sufficient evidence that the Negro American is genetically less endowed. Should anyone choose to apply in a practical fashion the data obtained thus far on race differences in IQ, the procedure would be extremely inaccurate. From a consideration of the problems of overlapping distributions and different "base rates" of Negroes and whites in the U.S. population (cf. Meehl & Rosen, 1955), it is possible to illustrate the practical futility of predicting race from a knowledge of IQ.

Let us use in our example the facts that 2 percent of the white standardization sample on the Binet obtain scores less than IQ 70 as contrasted with the 18 percent reported for the large representative sample of Southeastern Negro elementary schoolchildren described earlier. There are approximately 180 million whites and 20 million so-called Negroes in the United States at this time. If we choose to blindly label all individuals with tested IQs under 70 as Negro, the consequences are as follows: 3.6 million Negroes are accurately classified as to their race but 3.6 million whites are misclassified as Negroes. In the United States, using IQ under 70 as a criterion, you would be wrong 50 percent of the time if you were to use IQ as an indicator of race. You would be wrong more frequently than this if you were to use a higher cutting score such as IQ 80. With this score you could accurately identify 10.4 million Negroes but you would also label 14.4 million whites as members of the Negro race. Inasmuch as an individual's IQ does not permit you to accurately identify his race, so also his race does not permit you to estimate with sufficient accuracy his intelligence.

Let me summarize for you my answers to anyone's assertion that the 10 to 20 IQ point differentials observed between Negro groups and white groups represent evidence for genetic inferiority. Such differences in test scores cannot be equated with innate potential without an erroneous inferential leap—innate intellectual ability cannot be measured.

I find the usual evidence cited and especially the book by Audrey Shuey, The Testing of Negro Intelligence, to be anything but com-

pelling; not one single study cited by Shuey among 380 qualifies as a genetic analysis. The dreary litany fully documents that blacks obtain lower test scores than whites; such evidence, however multiplied, does not speak to the issue of innate differences, but may well attest to the effects on scholarship of slum housing, poor pre- and post-natal nutrition, low achievement motivation, and second-rate schooling.

1. We cannot accurately assess intellectual capacity or race; we don't know the neurophysiological bases for learning or the environmental variables that facilitate it.

2. It requires two niches, each with its own natural selection pressure, for a trait such as intelligence to show a genetic divergence between two races. Only 300 generations have elapsed since the option of agriculture versus hunting was available—not long enough nor a large enough difference in the demand characteristics of the ecologies to exert differential selection pressures for learning ability or problem solving in "Europeans" versus "Africans." (cf. Gottesman, 1968 for details).

3. The IQ differences observed do not compel a genetic explanation because they can also be explained otherwise. Identical twins have identical genes so that differences between their scores must be accounted for by post-conception differences in their environments; pairs of identical twins are known that differ by 35 points when one is brain injured, a grossly obvious example, and at least as much as 24 points when separated early and reared in fairly different households. Average differences between parents and their children or between brothers come to 12 points; we wouldn't think of assigning them to an innately inferior race. Knowing only that a man scored 12 fewer IQ points would not permit you to deny he was your brother.

4. I am impressed by data reported on longevity and hope there is a valid analogy to be drawn. Life expectancy for a white male in 1900 was 50-percent greater than for a Negro; by 1958 it was only 10-percent greater (67 versus 61). Such a fantastic change in so vital a trait can hardly be attributed to genetic changes occurring in a 58 year span. Meanwhile, the only humane and rational enterprise is to maximize the quality of the environment for every human. In my opinion the effect of maximizing the quality of the environment would be to shift the curve you saw for the Negro population with a mean IQ of 81 to the right so that the curve was more or less identical with the curve of the white population. This would mean then, that IQs under 100 would not be eliminated but that they would be equalized in the two populations.

I would now like to consider social class as a biological phenomena. You have just heard me support the view that race differences in intellectual ability are largely due to environmental factors. I would now like to make a case to support the view that social class differences in intelligence are in the process of becoming largely genetically based. It should be possible to examine the merits of such a position without subscribing to Social Darwinism or to the sickness of race and class prejudice.

The formation of social classes parallels the formation of races from a genetic point of view because both can be perceived as relatively isolated reproducing gene pools. I have just defined for you what the

population geneticist calls a Mendelian population, that is, a relatively endogamous breeding population.

Support for the view that the structure of modern societies is at least in part dependent on biological phenomena rests on the demonstration that societal stratification is based on ability and, further, that individual differences in ability are partially genetically conditioned. In a truly democratic system an open-class society permits the formation of differentiated social classes and, most importantly, fosters class change and social mobility. Thus a migration from one class to another based on the presence or the absence of ability is the final essential requirement for a biologically based model of social structure.

Wherever class differences and intelligence, as measured by IQ tests, have been examined, a spread between the average score of the highest and the average score of the lowest class has been found. You can see this phenomenon clearly in the next slide. The data were collected by Cyril Burt in the greater London area during the period 1913–1960. He looked at the intelligence scores of some 40,000 adults and their children as a function of occupational status. The sample sizes have been made proportional to a base of 1,000. Notice that the mean IQs for adults range from about 140 in social class 1, the higher professionals, down to an IQ of about 85 for unskilled workers in social class 6. Notice also that there is quite a bit of variation with respect to IQ score within each of the occupational classes. Even the unskilled class has individuals with IQs going up to 120.

The next slide shows the same kind of information, but for the children of the men who occupied the various occupational classes in the last slide. The first thing to notice in this slide is that the children have a much narrower range of IQ score averages than did their fathers. The children of the highest occupational class have a mean IQ of about 121 compared to their father's score of 140, whereas the children of the unskilled workers have a mean IQ of about 93 compared to their father's score of 85. I should add that similar data have been collected in the United States with similar results.

It is apparent that the mean intelligence score of the children in each class is closer to the population mean of 100 than their fathers' and the IQs of the children vary much more than their fathers. Within the occupational classes the standard deviation for adults is 9.6 contrasted with a standard deviation of 14 for their children. You may recall that the value of the standard deviation for IQ test scores in the total population is 15 points.

Evidence now available shows a constant gradient of high to low IQ scores for the occupational distribution of IQs from one generation to the next generation, for example from WW I to WW II. This must mean, therefore, that if the children shown in this table are to have the same distribution of IQ scores when they are adults as their fathers, a large number will have to migrate to a social class different from that of their fathers. If, for purposes of illustration, we assume that vocational adaptation depends exclusively on intelligence as measured by these tests, many adults have too much or too little intelligence for their roles in life. Cyril Burt calculated that only 55 percent of the adults in his British sample were correctly placed; the rest were either above

their level or below their level and were square pegs in round holes, so to speak.

The greater variation among the children of an occupational class would make for even greater mismatching between ability and role if there were no social mobility. In order to estimate the extent of social mobility for the British population, the task became one of estimating the changes required to bring the distribution for the children in line with that of the adults. When this was done, it turned out to lead to an estimate of intergenerational social mobility in the neighborhood of 30 percent. That means that 30 percent of children move above or below their own familes' social status.

In comparison to the social structure of Great Britain, the open-class aspects of democratic society, at least for the white-skinned American, are much more pronounced in the United States. In this decade approximately 50 percent of high school graduates are enrolling in college, and social mobility is probably greater than 30 percent per generation.

The next slide shows the results of a study just completed by Dr. Jerome Waller, a student from our doctoral training program in behavioral genetics at the University of Minnesota. His data, based on 174 father-son pairs, show quite clearly that the amount and direction of social mobility is connected strongly with the difference in intelligence between a father and his son. The greater the difference in intelligence the greater is the probability of a change in social status and the greater is the number of steps skipped in the social structure. You can see from the figure that of those sons who were 15 or more points higher than their fathers', more than 60 percent improved their social position; of those sons who were 15 or more IQ points lower than their fathers', more than 70 percent suffered a loss in their social position. These kinds of data may also account, in part, for the so-called generation gap.

The picture I am trying to draw for you is a statistical one, in fact the correlation between IQ and occupational level is only 0.5. I do not mean to imply that all of the genetically controlled constitutional factors responsible for high achievement and intelligence are confined to the highest social layer or that all of the factors responsible for nonachievement are to be found in the lowest layer. It is obvious that all factors are found in all social layers. The layers differ, however, with respect to the relative frequency with which the factors responsible for high and low achievement occur. The existence of class barriers, however permeable they may be, fosters relative reproductive isolation; yet social mobility permits a constant winnowing for achievement and learning ability. Migration to an appropriate social ecological niche follows. The net result of an open-class democratic system together with equality of opportunity and assortative mating, like marrying like, is to make genetic factors no less important for an understanding of the structure of human society than they are for understanding other mammalian species.

Let me digress a moment to support a point I may have made too casually. Individual differences within a group exposed to a more or less homogeneous environment are due to genetic factors. The next slide shows the degree of similarity in IQ test scores for different pairs of relatives varying in their degree of genetic similarity. You can see that identical twins, with all their genes in common, are most like each other, even when they are raised apart from one another. Fraternal

twins, ordinary siblings, and parents and their children, all have half their genes in common, on average and the degree of similarity in score drops accordingly to a correlation of 0.5. And so forth for more remote relatives down to unrelated persons paired at random with a correlation of zero.

The next slide is meant to show you that IQ is not caught like measles from the people around you. The data are from a study by Skodak and Skeels (after Honzik, 1957) showing the resemblance over time, from age 2 to age 14, between foster children's IQ and their foster mothers' educational level in the lower-line. Correlations over time stay very close to zero; in other words, there is very little, if any, relationship between the educational level of a foster mother and the IQ of children not related to her biologically but whom she is raising. Now look at the higher-line. This shows the degree of similarity between the same foster children and their biological mothers' education. These children have had negligible contact with their biological mothers. Despite that fact, the similarity between them grows in the same fashion as that in a control group of children reared by their own biological mothers. Those data are shown in the higher solid line. The two top curves end up with the correlation close to 0.4. In the preceding slide I showed you that the relationship usually found between parents and their children with respect to IQ is about 0.5.

CONCLUSION

I will leave to Dr. Lourie the burden of discussing the social relevance and the implications of my remarks for social work practice. There has been much talk recently about the concept of heritability of intelligence (Jensen, 1969a,b; Hirsch, 1970). The heritability of IQ test scores is in fact very high (Gottesman, 1963; Vandenberg, 1968) in the neighborhood of 80 percent. However, high or low heritability of human traits tells us nothing practical about how a particular individual might have developed under conditions different from those in which the experiments giving us our heritability values were conducted. Heritability is a characteristic of a trait in a specific population or gene pool; it provides no useful information about the reaction range of a particular genotype.

I would like to enter some data into the record although I have not read the original report of them. The data were cited in passing by Prof. Benjamin Bloom of the University of Illinois (1969). The amount and intensity of social experimentation in Israel is much greater than it is in our own country. The Israeli experience with Kibbutz childrearing demonstrates that while Jewish children of European origin normally have an IQ of about 105 under home-rearing conditions, under Kibbutz nursery rearing of 22 hours a day for 4 or more years, they have an average IQ of 115. Middle Eastern Jewish children, however, have an average IQ of 85 under their own home-rearing conditions. Under the intensive nursery-rearing regime, they have an average IQ of 115. The two groups of children were matched with respect to the educational and occupational level of their parents as well as the particular Kibbutz in which they were reared. To me, these kinds of data illustrate beautifully the concepts of genotype by environment interaction and of the reaction range of genotypes.

Although my role here today has been that of an emissary from the scientific establishment, I would not want you to perceive me as an apologist for the status quo. The status quo stinks. Social ethics, justice, and equality of rights must be independent of empirical or scientific "facts." At the same time, our concern for equality of rights must not lead us into a trap where we are forced to deny the existence of genetically determined individual differences; equality does not equal identity. As my friend Brendan Maher (1969) has noted, "Equality of rights is a moral axiom that cannot be upset by any set of scientific data. Among these rights is the right to an education that will develop the potential of every individual. The obligation to continue to improve teaching techniques does not depend upon a resolution of the nature-nurture controversy in favor of nurture, it rests upon the belief that our present techniques fall so far short of optimal that there is little reason to suspect that we are providing a better education than some children can use."

TABLE 1.—MEAN INTRAPAIR DIFFERENCES IN IQ TEST SCORES

Study	N	Test	Mean differences in IQ points	Range of differences in IQ points
Newman, Freeman & Holzinger (1937)	19	Stanford-Binet	8.2	1–24
Shields (1962)	38	Combined Score	9.5	0–30
Shields (1962) as reported by Shields & Gottesman.	38	Mill Hill Vocabulary	14.0	0–23
		Dominoes	10.0	
Juel-Nielsen (1964)	12	Wechsler-Bellevue	7.3	1–14
Burt (1966)	53	Stanford-Binet (London standardization).	6.0	

TABLE 2.—DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: ADULTS (FROM BURT, 1961)

IQ	Professional: Higher I	Professional: Lower II	Clerical III	Skilled IV	Semiskilled V	Unskilled VI	Total
50 to 60						1	1
60 to 70					5	18	[illegible]
70 to 80				2	15	52	[illegible]
80 to 90			1	11	31	[illegible]	[illegible]
90 to 100			8	51	135	53	247
100 to 110			16	101	120	11	[illegible]
110 to 120		2	56	78	17	9	162
120 to 130		13	38	14	2		67
130 to 140	2	15	3	1			21
140+	1	1					2
Total	3	31	122	258	325	261	1,000
Mean IQ	139.7	130.6	115.9	108.2	97.8	84.9	100

TABLE 3.—DISTRIBUTION OF INTELLIGENCE ACCORDING TO FATHER'S OCCUPATIONAL CLASS: CHILDREN (FROM BURT, 1961)

IQ	Professional: Higher I	Professional: Lower II	Clerical III	Skilled IV	Semiskilled V	Unskilled VI	Total
50 to 60					[illegible]	[illegible]	[illegible]
60 to 70				[illegible]	[illegible]	[illegible]	[illegible]
70 to 80			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
80 to 90		1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
90 to 100		2	21	[illegible]	[illegible]	[illegible]	[illegible]
100 to 110	1	[illegible]	31	[illegible]	[illegible]	[illegible]	[illegible]
110 to 120		12	[illegible]	[illegible]	[illegible]	16	[illegible]
120 to 130	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
130 to 140	1	[illegible]	[illegible]	[illegible]	[illegible]		21
140 plus	[illegible]			1			1
Total	3	31	122	258	325	261	1,000
Mean IQ	120.8	114.7	107.8	104.6	98.9	92.6	100

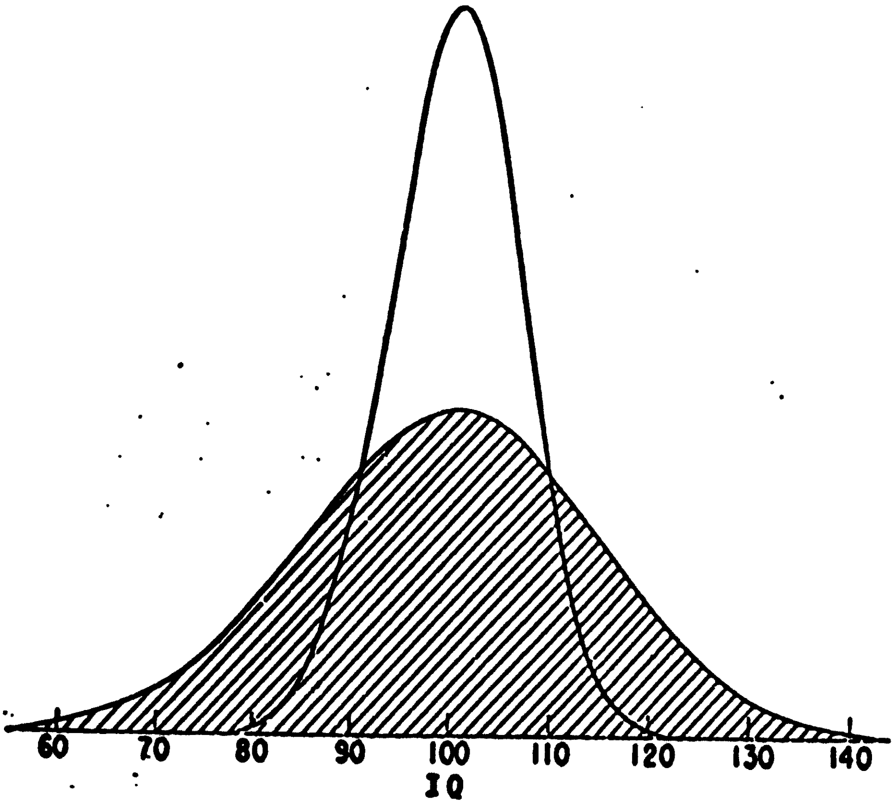


FIGURE 1

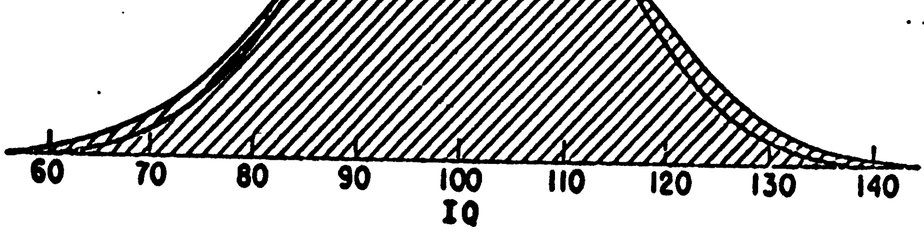
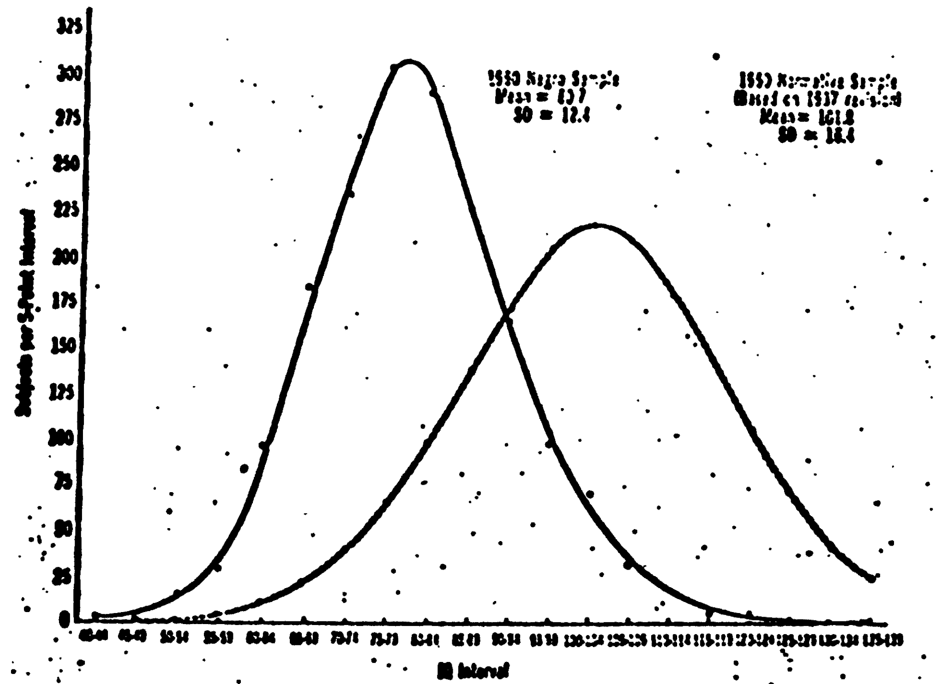


FIGURE 2



The IQ distributions of normative white and southeastern Negro school children. (After Kennedy *et al.*, 1963)

FIGURE 8

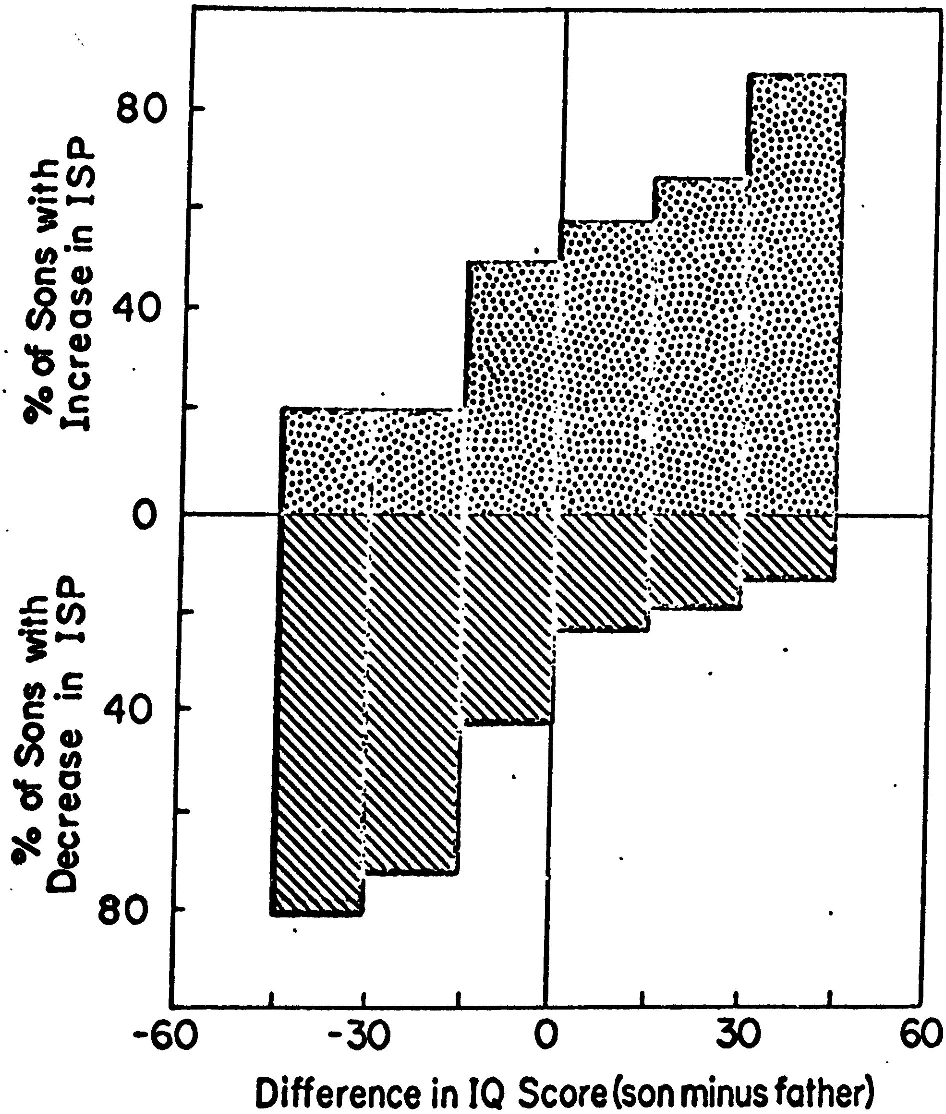


FIGURE 4.—Relationship of difference in social position to difference in IQ score.

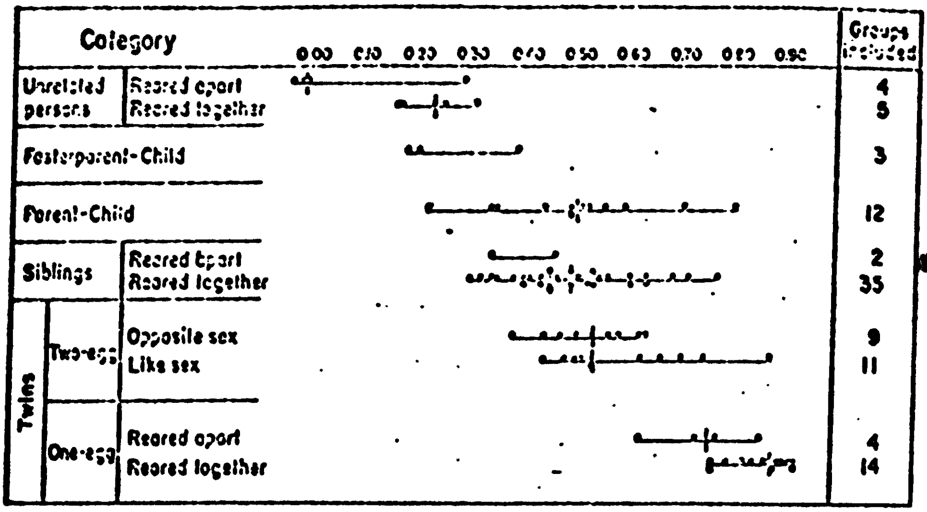


FIGURE 5.—Correlations between IQ's of paired individuals of genetic relations ranging from none to complete. Dots represent correlations from single studies; lines show range of values; median is shown by short vertical lines. (Erlenmeyer-Kimling and Jarvik, 1963)

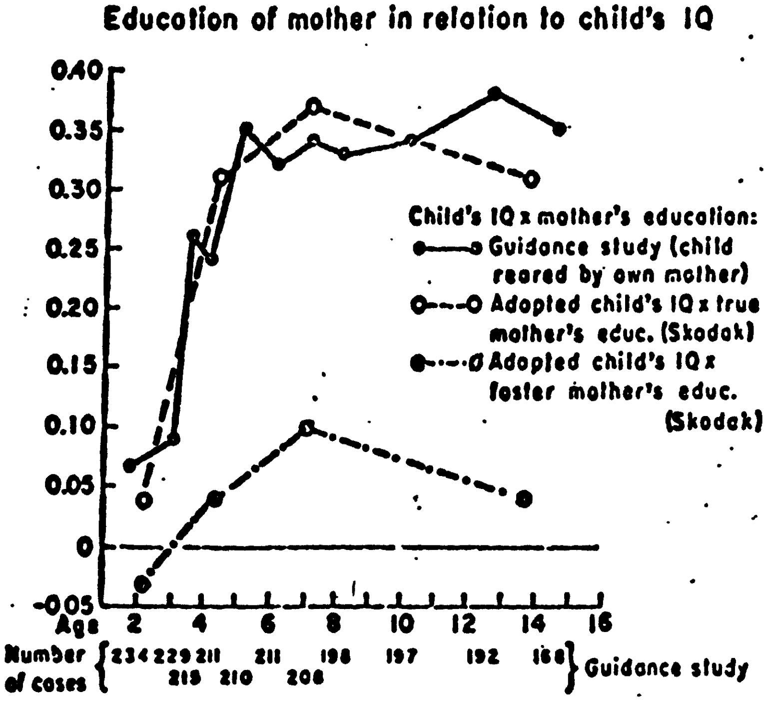


FIGURE 7.—IQ resemblance of adopted child to foster and true parents. (After Honzik, 1957)

BIBLIOGRAPHY

Bloom, B. Letter to the editor. Harvard Educational Review, 1969, 39, 419–421.

Burt, C. Intelligence and social mobility. British Journal of Statistical Psychology, 1961, 14, 3–24.

Bressler, M. Sociology, biology and ideology. In D. C. Glass (Ed.) Biology and Behavior: Genetics. New York: Rockefeller University Press, 1968, 178–210.

Cravioto, J. Nutritional deficiencies and mental performance in childhood. In D. C. Glass (Ed.) Biology and Behavior: Environmental influences. New York: Rockefeller University Press, 1968, 3–51.

Duetsch, M. Happenings on the way back to the Forum. Harvard Educational Review, 1969, 39, 523–557.

Dunn, L. C. Cross currents in the history of human genetics. American Journal Human Genetics, 1962, 14, 1–13.

Erlenmeyer-Kimling, L. and Jarvik, L. F. Genetics and intelligence: a review. Science, 1963, 142, 1477–1479.

Glass, D. C. (Ed) Biology and Behavior: Environmental Influences. New York: Rockefeller University Press, 1968.

Gottesman, I. I. Genetic aspects of intelligent behavior. In N. Ellis (Ed.), Handbook of mental deficiency: psychological theory and research. New York: McGraw-Hill, 1963, 253–296.

Gottesman, I. I. Personality and natural selection. In S. G. Vandenberg (Ed.) Methods and goals in human behavior genetics. New York: Academic Press, 1965, 63–80.

Gottesman, I. I. Biogenetics of race and class. In M. Deutsch, I. Katz and A. Jensen (Eds.) Social class, race and pyschological development. New York: Holt, Rinehart & Winston, 1968, 11–51.

Gottesman, I. I. Testing Negro intelligence: comments on Eysenck. The Humanist, 1970, 30, 35–36.

Haller, M. H. Eugenics: hereditarian attitudes in American thought. Brunswick: Rutgers University Press, 1963.

Hirsch, J. Behavior-genetic analysis and its biosocial consequences. Seminars in Psychiatry, 1970, 2, 89–105.

Honzik, Marjorie P. Developmental studies in parent-child resemblance in intelligence. Child Development, 1957, 28, 215–228.

Hunt, J. McV. Intelligence and experience. New York: Ronld, 1961.

Hunt, J. McV. Environment, development and scholastic achievement. In M. Deutsch, I. Katz and A. Jensen (Eds.) Social class, race and psychological development. New York: Holt, Rinehart & Winston, 1968, 293–336.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969(a), 39, 1–123.

Jensen, A. R. Reducing the heredity-environment uncertainty: a reply. Harvard Educational Review, 1969(b), 39, 449–483.

Maher, B. A. Intelligence and brain damage. In N. Ellis (Ed.) Handbook of mental deficiency: psychological theory and research. New York: McGraw-Hill, 1963, 224–252.

Maher, B. A. Intelligence, biology, and politics. Psychology in the Schools, 1969, 6, 152–155.

Meehl, P. and A. Rosen. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 1955, 52, 194–216.

Newman, H., F. Freeman, and K. Holzinger. Twins: A study of heredity and environment. Chicago: University of Chicago Press, 1937.

Shields, J. Monozygotic twins brought up apart and brought up together. London: Oxford University Press, 1962.

Shuey, Audrey M., The testing of Negro intelligence (second edit.) New York: Social Science Press, 1966.

Thoday, J. M. and Gibson, J. B. Environmental and genetical contributions to class difference: a model experiment. Science, 1970, 167, 990–992.

Vendenberg, S. G. The nature and nurture of intelligence. In D. C. Glass (Ed.) Biology and behavior: Genetics. New York: Russell Sage, 1968, 3–58.

Waller, J. H. The relationship of fertility, generation length, and social mobility to intelligence test scores, socioeconomic status, and educational attainment. Unpublished doctoral dissertation. University of Minnesota, 1969.

STATEMENT OF ARTHUR R. JENSEN

I—Introduction

Mr. Chairman and members of the committee, my name is Arthur R. Jensen. I am Professor of Educational Psychology at the University of California at Berkeley, where I have been a member of the faculty for the past 14 years.

First, I want to thank this committee for inviting me to summarize my position on the topics mentioned in your invitation: "heredity, environment, intelligence, and scholastic achievement." My position on some aspects of these questions, as you know, is a controversial and unpopular one; but, it is a position which I have come to through my intensive studies, over the past decade, of virtually all the relevant evidence in this field and through my own experimental and statistical investigations. The reporting of these conclusions in scholarly journals, as in my well known article in the Winter 1969 issue of the *Harvard Educational Review*,* has not been without some penalty. At least I have heard of no one taking the contrary position who, as a result, has had to endure personal threats, prolonged harassment, invasion of his college classes by disruptive demonstrations, or has had his university administration take such precautions as putting two plainclothes police bodyguards on him—in lectures, on the campus, in going to and from classes and to the parking lot. Much of this emotional reaction to my *Harvard Educational Review* article, I believe, is a result of the fact that a generation of social scientists and educators has been assiduously indoctrinated to believe that genetic factors are of little or no importance in human behavior and human differences. My summaries of the relevant evidence contradict this belief. The reactions have often been extreme. In a recent Newsletter of the American Anthropological Association, for example, it was proposed, apparently in all seriousness, that members of the AAA should burn—literally burn—all copies of the *Harvard Educational Review* containing my article! I am therefore grateful for this opportunity to make clear my position on these important matters to the committee; and also, to see that a number of distinguished scientists are concerned with some of the issues I have raised and are here today to express their views. Though I do not expect all their opinions to coincide completely with mine—indeed, they were selected expressly to insure that a variety of views might be expressed before this committee—I am especially gratified to see that there are prominent scientists who, like myself, are trying to come to grips with these difficult questions now under discussion.

*See page 710.

II—Scholastic Achievement and Intelligence

Scholastic achievement is what children learn in school—a repertoire of knowledge and skills, including the 3 R's. Scholastic achievement can be most reliably measured by means of standardized tests at every stage of schooling from kindergarten to college. When entire school populations are tested on scholastic achievement, there is revealed a wide range of individual and group differences at any given grade level. In one 5th grade classroom in San Francisco, for example, achievement levels ranged from 1st to 11th grade (*Chronicle*, Feb. 21, 1972). What most educators, government officials, and writers in the popular press who talk about the present problems of education are, in fact, referring to is not primarily dissatisfaction with some *absolute* level of achievement, but rather with the large group *differences* in educational attainments that show up so conspicuously in our educational system—the achievement gaps between the affluent and the poor, the lower-class and the middle-class, and the majority and the minority, the urban and the suburban, and so on. Educational *differences*, not absolute level of performance, is the main cause of concern. Whether we like to admit it or not, the problem of achievement differences today is where the action is—where the billions of dollars of educational funds are being poured in, where the heat is on, and where the schools are being torn apart. We are trying to understand more about the causes of these differences. Massive surveys and statistical analyses such as James Coleman's* well-known report on *Equality of Educational Opportunity* (1966) have shown that only a small fraction (about one-tenth to one-fifth) of the total variation in scholastic achievement is attributable to factors in the schools themselves. In other words, differences among schools and school systems nationwide are not sufficiently large to account for more than a small fraction of the total variation in scholastic achievement at any grade level.

The single most powerful predictor of children's scholastic performance is *intelligence* as measured by any one of a variety of standard intelligence tests: group or individual tests, verbal or nonverbal. No other variable has yet been found which makes as large an independent contribution to variance in scholastic achievement as does intelligence. I have found in my own research that a composite of several different measures of intelligence will predict nearly all the true variance in scholastic achievement scores, and the predictive validity of the intelligence tests becomes better with advancing grade level. Furthermore, the validity of intelligence tests for predicting scholastic achievement is not significantly different for white and black children. In this respect, intelligence tests are quite color blind. That is to say, a white child and a black child with the same IQ can be expected to perform about equally well in school. In short, if intelligence tests can be said to be good for anything, they are good for predicting scholastic achievement.

Intelligence, in the technical sense that psychologists use this term, is not the same thing as scholastic achievement. Schools do not teach

---

*See also, Select Committee on Equal Educational Opportunity hearing of April 21, 1970, Part 1A—Equality of Educational Opportunity, An Introduction; pp. 87–184.

intelligence per se. Intelligence is mental brightness; it is a capacity for conceptualization, abstract reasoning and problem solving, for processing information in the form of words and symbols, for integrating and understanding what is learned, and for making broad transfer from past learning to the solution of novel problems. As Harvard psychologist Professor Lawrence Kohlberg recently noted, scholastic achievement merely rides on the back of intelligence: ". . . bright kids learn the stuff they're taught in school faster, but learning the stuff they're taught in school doesn't make them brighter."*

Intelligence *can* be quite reliably measured by appropriate tests, and these measurements, often called IQs, show substantial correlations with a number of educationally, occupationally, and socially important criteria. The correlations are not appreciably different for children of the poor, for whites or for blacks. Although intelligence—as psychologists use this term—is the single most powerful, though by no means perfect, predictor of academic performance, it is surely not the whole of mental ability and human competence. To equate intelligence with all virtue is a ridiculous mistake. But in discussing achievement differences in schools with their present curricula and their present instructional methods, intelligence differences, we know, are of central importance. There are undoubtedly other socially important mental abilities besides intelligence, and on some of these abilities we find little or none of the social class and racial differences that we find for intelligence. Much of my own research has been concerned with identifying such abilities and with trying to determine their relevance to instructional methods that might make better use of these other abilities for scholastic learning. I have summarized some of this research in *A Two-Factor Theory of Familial Mental Retardation.***

Both intelligence and scholastic achievement grow or increase in a quite regular manner in most children and their individual mental growth curves become increasingly stable over the years from infancy to maturity.

III. The Heritability of Intelligence

A number of genetic studies carried out in Europe and the United States over the past 40 years provided evidence which shows quite conclusively that, in the populations studied, a very substantial proportion of the variability among persons in intelligence is attributable to genetic, i.e., inherited, factors. The vast majority of studies have found that the proportion of population variance attributable to genetic differences is something between the extreme limits of about .60 to .90, a figure known in the broad sense as the *heritability*. Since heritability—i.e., the proportion of the total variance which is genetically determined—is a population statistic subject to sampling error and other sources of variation, it has no universal or constant value for all times, all tests, and all populations. But empirically determined values are usually of the order of .70 to .80. As I explain in more detail in *The Heritability of Intelligence*,*** I believe it is safe to

**Education Summary*, August 6, 1971.
**See page 195.
***See page 214.

say that in European and North American Caucasian populations, at the present time, genetic or hereditary factors are roughly twice as important as environmental variation as a cause of individual differences in intelligence as assessed by standard tests.

Another way of stating this is in terms of what geneticists call the *reaction range* for the phenotypic expression of intelligence. This is the range through which intelligence can vary due to nongenetic influences. The total reaction range of IQ in those populations in which it has been determined—e.g., London school children—is about 25 to 30 IQ points. That is to say, two individuals with exactly the same genetic make-up—e.g., monozygotic twins—could differ by as much as 25 to 30 points if one, from the moment of conception, had been subjected to the lowest—i.e., least favorable—1 percent or less of all those environmental factors, prenatal and postnatal, involved in mental development; and, the other had been subjected to the better than the highest—i.e., most favorable—1 percent of environmental factors. Such extreme environmental differences theoretically should result in about a 25 to 30 point IQ difference even when there is no genetic difference between the individuals. This conclusion, however, does not necessarily mean that we know what all of the nongenetic factors are that are involved in this wide reaction range or that we can control or manipulate all of the nongenetic factors involved. The extremely important experiment of Professor Rick Heber—which you will hear about later on—illustrates, among other things, this reaction range concept of mental development. He has compared children reared in what may well be the lowest 1 or 2 percent of environmental conditions found in our society with genetically similar children reared in the very most mentally stimulating environments that we know how to devise and is probably beyond the scale of naturally occurring environments, and he finds IQ differences consistent with those predictable from our knowledge of the heritability and the reaction range of IQ.

Still another way of stating this is that one standard deviation increase in the totality of environmental factors that account for all the nongenetic variance in IQ will increase IQ about 5 points. In order to explain the 15 points, average, white-black IQ difference solely in terms of these environmental effects, it would have to be assumed that there is almost no overlap between the environmental conditions of the black and white populations. In what aspects of the environment can one point to virtually nonoverlapping differences between blacks and whites? Of course, the demonstration of any environmental difference does not in itself prove that this difference is the *cause* of intelligence differences. The observation of an environmental difference at best suggests an hypothesis which must then be tested in its own right.

The evidence on the heritability of intelligence in the American black population is necessarily much more tentative than the evidence in white populations, since there are only two reasonably good studies of the matter—by Sandra Scarr-Salapatek and by Paul L. Nichols—and even these present some very serious statistical and methodological ambiguities which, at present, rule out any firm conclusions. From these studies it appears that the heritability of IQ in the black population may be about the same or slightly lower than in the white. As statistical tests of significance were not performed by these investigators,

to say much more at this time would be sheer speculation. A more detailed commentary on the one of these studies which is now published is herewith submitted as *Technical Comment on Scarr-Salapatek's "Race, Social Class, and IQ"**.

IV. Socioeconomic Status and IQ

When persons are classified into several socioeconomic status (SES) categories according to their occupation, education, and income, the mean IQ of each of these groups differ from one another, with the highest SES group—professional and managerial—having the highest mean IQ and the lowest SES group—unskilled labor—having the lowest IQ, with an average IQ difference between these extreme groups of at least 40 points. The spread of IQs within each group, however, is considerable. The children born into these SES categories also differ, on the average, but not as much; the average difference between the highest and lowest categories is about 20 to 30 points. Within each SES group, however, the spread of the children's abilities is nearly as great as in the total population. All levels of intelligence are found in every social class, although their frequencies differ.

The evidence, in my opinion, is now quite substantial that genetic factors contribute strongly to these average social class IQ differences within racial groups. I have found no disagreement with this conclusion among geneticists who have studied the evidence. These average IQ differences between social classes are attributable in large part to genetic differences that result from differential selection of the parent generations for different patterns of ability. Those abilities most highly related to educational achievement make for social mobility in our society and this results in some degree of genetic difference between social classes in average potential for intellectual development and in scholastic achievement.

V. Race Differences in IQ and Achievement

In the United States the average IQ difference between blacks and whites is approximately 1 standard deviation—i.e., 15 IQ points—which is about one-sixth of the total range of IQs found in 99.9 percent of the population. The average white-black difference is slightly more or less than 15 points in different regions of the country. There is a great spread of IQs in both the white and the black populations, and some 12 to 15 percent of blacks exceed the average white on IQ tests. The one standard deviation mean difference is also statistically consistent with the approximately six times higher rate of severe educational retardation, with IQs below 70, in the black population as compared with the white.

What are the causes of this IQ difference? *Here let me make my position perfectly clear.* The storm of criticisms that have been leveled at me has been a result of my expressing serious *doubts* that this racial difference is entirely explainable in terms of culture-bias in tests, unequal education opportunities, social discrimination, and other environmental influences. My position is that there is now sufficient evidence

---

*See page 218.

to seriously question the 100-percent environmental theories of the mean white-black intelligence difference. Are there any responsible scientists today who claim that this position can be ruled out on the basis of evidence or ruled out *a priori* by any principle of genetics? How many scientists today express little or no doubt that all of the racial IQ difference is attributable to environment; and, on what evidence do those who claim no doubt base their certainty? I have not found any 100-percent theory which can explain the facts or which stands up when its major premises are critically examined in the light of evidence. Therefore, I regard this as an open question which can be eventually answered in a scientific sense only if we are willing to consider all reasonable hypotheses. It is a reasonable hypothesis that genetic factors are involved in the average white-black IQ difference, and my study of the research evidence is most consistent with a genetic hypothesis, which, of course, does not exclude the influence of environment as well. My position on this point is elaborated more fully in *Can We and Should We Study Race Differences**. An hypothesis that I believe comprehends more of the facts and is consistent with more of the converging lines of evidence than any other I know of, in its simplest terms, is the hypothesis that:

1. The heritability of IQ is the same *within* the white and black populations as *between* the populations; and,
2. The genetic variance involved in IQ is about one-fifth less in the black than in the white population.

This, I repeat, is an *hypothesis*. It is also, in my opinion, more consistent with all of the evidence I have reviewed than any other hypotheses I have seen presented.

Studies which estimate the heritability of intelligence *within* each of two populations, such as blacks and whites, do not provide any formal proof that an observed average difference between the populations is attributable to genetic factors. Other methods than heritability analysis are required to determine the degree of genetic difference between two populations. However, high heritability of a trait *within* populations that differ in the trait, as whites and blacks differ in intelligence, does increase the *a priori* likelihood that there exists a genetic difference between the populations. The fact of the high heritability of IQ, therefore, makes it a very reasonable and likely hypothesis that genetic factors are involved in the white-black IQ difference. No geneticist to my knowledge has argued the contrary.

The black-white difference in scholastic achievement, contrary to popular opinion, is slightly less than the difference in IQ, especially if IQ is measured by nonverbal tests that have very little resemblance to tests of scholastic achievement. More detailed information on this point is given in *Do Schools Cheat Minority Children?***.

VI. Environmental Theories

A genetic hypothesis does not rule out environmental factors. In this sense it is *not* the diametrical *opposite* of environmentalist theories, which absolutely exclude any question of genetic differences. Some social scientists have tried to explain away racial differences by

---

*See page 223.
**See page 245.

statistically equating racial groups for socioeconomic status (SES) and related environmental variables in order to show a smaller IQ difference. This is what I have termed the "sociologist's fallacy"; it assumes that all the environmental variables on which the groups were statistically equated are direct *causal* factors, when in fact they are merely *correlates* of IQ. If some part of the SES difference within racial groups has a genetic basis, then statistically equating racial groups on social class equates them also to some degree on the genetic factors involved in intelligence. Indeed, it is theoretically conceivable that if one equated racial groups on a large enough number of *correlates* of IQ, one could statistically eliminate all of the IQ difference between them. But it would prove nothing at all about the *causes* of the mean IQ difference between the total populations. Many environmental indices are undoubtedly correlated with genotypes. Educational level of the parents, for example, is often included as an environmental variable affecting the child's development. But it almost certainly includes also some genetic component which is common to both the parents and their children. If the environmental variables used for statistical control account for more of the IQ variance *within* racial groups than the complement of the heritability—i.e., $1-h^2$—within the groups, then it is virtually certain that the environmental indices also reflect correlated genetic factors.

I have carefully examined the major environmentalist theories of the black-white IQ differences, and I find that none of them stands up in the face of the existing relevant evidence. I have reviewed this evidence in detail in *Educability and Group Differences*, which is a chapter from my forthcoming book. I summarize some of these points below.

Those environmentalist hypotheses of the Negro-white IQ difference which have been most clearly formulated and are therefore subject to empirical tests are the only ones that can be evaluated within a scientific framework. The most frequently cited environmentalist hypotheses which are sufficiently clear to put to an empirical test and which already have been put to a test have not proven adequate to the explanatory function they were intended to serve. A number of lines of such evidence cast serious doubt on purely environmental and cultural theories of the racial IQ difference.

1. Socioeconomic Differences

Matching or statistically controlling the SES of racial samples does not wipe out IQ differences, although on the average it reduces differences by about one-third. The racial difference increases with increasing SES level when the IQs being compared are those of children classified according to their parents' SES. The finding is hard to rationalize along purely environmental lines, but it is predictable from the genetic principle of filial regression toward the population mean. Children of high SES Negro children regress toward the lower mean of the Negro population, which makes for a larger regression effect for white children, who regress toward the white population mean—which is about one *SD* higher. High SES Negro parents are more deviant from their population mean than high SES white parents, and there-

fore, because of regression, the Negro children will resemble their parents less in ability than will the white children.

Criticisms of studies that control for SES argue that SES—as it is usually measured in terms of parental education, occupation, income, etc.—is too crude a variable to reflect adequately the important variables of the environment influencing mental development. But SES is crude mainly in the sense that it is nonspecific; it summarizes within it many other environmental variables, and adding more refined variables does not markedly increase a multiple correlation with either IQ or with race. SES seems to summarize the larger part of those environmental factors that are most frequently mentioned as the causes of racial IQ differences. Actually, rather than controlling too little of the variance, SES probably controls too much, since within racial groups, at least, there is undoubtedly a correlation between SES and genetic factors. Matching racial groups for SES thus matches them not only for some environmental factors but also to some unknown extent for genetic factors as well. It is interesting also that when such matching is carried out, it is noted that the average skin color of the Negro groups becomes lighter in the higher SES categories, indicating that genetic factors covary with SES, for whatever reason. Genetic SES intelligence differences are firmly established within the white population. Matching Negro and white groups on SES, therefore, is certain to minimize genetic as well as environmental differences. For this reason, studies that control for SES are probably biased in favor of the environmentalist hypothesis and can contribute little or nothing to elucidating the nature-nurture problem, except in those instances where the direction of the environmental difference between two groups is opposite to the direction of the IQ difference.

2. Negative Correlations Between Environment and Ability

A number of environmental factors which correlate positively with mental ability *within* various population groups have been shown to correlate *negatively* with IQ differences *between* certain groups. On all of the many measurable factors which environmentalists have invoked to explain the Negro-white IQ difference, both American Indians and Mexican-Americans have been found to be much more disadvantaged than Negroes. Yet on nonverbal intelligence tests—which are more fair for bilingual groups such as Mexicans and Indians—and in scholastic performance, Indians and Mexicans significantly outperform Negroes. This finding is neutral with respect to a genetic theory, in the sense that no prediction could have been derived from genetic principles; but it contradicts those environmental theories that invoke measurable environmental factors known to correlate with IQ within population groups as the cause of the lower Negro IQ. The only attempts of environmentalists to rationalize these findings have invoked highly speculative cultural and attitudinal factors which have not yet been shown to be correlated either with IQ or with race.

3. Culture-Based Tests

Intelligence tests can be rank-ordered according to certain generally agreed upon criteria of their cultural loading. Within a given culture,

tests are better described as differing in *status fairness*. Environmentalists who criticize intelligence tests usually give as examples those tests which are most obviously loaded with what is presumably white, middle-class factual knowledge, vocabulary, and the like, as contrasted with more abstract figural material such as compose Raven's Progressive Matrices and Cattell's Culture-Fair Tests of *g*. Yet it is on the latter type of tests that Negroes perform most poorly, relative to whites and other minority groups. Disadvantaged minorities, such as American Indians and Mexican-Americans, perform on tests showing different degrees of status bias in accord with the environmentalist hypothesis. Negroes do the opposite. "Translation" of tests such as the Stanford-Binet into the Negro ghetto dialect also does not appreciably improve scores.

The scholastic and occupational predictive validity of IQ tests is the same for Negroes as for whites, and item analyses of tests showing large average group mean differences do not reveal significant differences in rank order of item difficulty or in choice of distractors for error responses. Test-taking attitudes and motivational factors appear unconvincing as an explanation of the group difference in view of the fact that on some tests which make equal demands on attention, persistence, and effort, such as various memory tests, Negroes do perform quite well relative to whites. When various diverse tests and test items are ordered in terms of the degree to which they discriminate between Negroes and whites, the one feature which is common to the most discriminating tests and items is the conceptual and abstract nature of the test material, or the degree to which they accord with the classic definitions of the psychological nature of *g*. Data from other minority groups who are more environmentally disadvantaged than Negroes support an opposite conclusion, in accord with the environmental interpretation.

4. Language Deprivation

This is an unconvincing explanatory hypothesis in view of the fact that Negroes perform best on the most verbal parts of intelligence tests and poorest on the least verbal materials. All other disadvantaged minority groups within the American population show the opposite trend. Children who are born deaf are the most verbally deprived subjects we can study. They show marked verbal deficits on intelligence tests. Yet they perform at an average level on nonverbal tests, thus showing a pattern of abilities opposite to that of Negroes.

5. Poor Motivation

There is no consistent evidence that Negroes are less motivated in a test situation than are other groups. Some groups—e.g., Indians—whose general educational aspirations and self-concepts are poorer than those of Negroes actually perform better on tests and in school. Also, on performance tests specially devised to maximize the influence of motivational factors and to minimize the test's dependence upon abstract or complex cognitive functions which would involve *g*, Negroes do not perform significantly below whites. The "expectancy" or "self-fulfilling prophecy" theory has not been empirically demonstrated, and when put to proper tests it has failed to be substantiated.

6. Noncognitive Tests

Certain perceptual-motor tests such as choice reaction time and pursuit rotor learning—which has a very high heritability—show large Negro-white differences even under very highly controlled experimental conditions, and the results are independent of the race of the tester. Moreover, the magnitude of the racial difference has been shown to be related to the degree of Caucasian admixture in the Negro sample as assessed by physical indices. If genetic racial differences in behavioral tests other than intelligence tests are admitted, by what principle can one exclude the same possibility for types of tests labeled as measures of intelligence? There is no reason why intelligence tests should be categorically excluded from the possibility of showing genetic race differences when such differences in other physical and behavioral traits can be found.

7. Nutritional Deficiencies

The fact that severe malnutrition, especially protein deficiency, during prenatal development and in infancy and childhood can impair mental as well as physical growth is not at issue. Studies from the most nutritionally deprived segments of populations in Africa, Mexico, and South America would support this conclusion. There are no data, however, which would support the hypothesis that malnutrition contributes any appreciable fraction to the average Negro-white IQ difference. In Negro communities where there is no evidence of poor nutrition, the average Negro IQ is still about one *SD* below the white *mean*. When groups of Negro children with IQs *below* the general Negro average have been studied for nutritional status, no signs of malnutrition have been found. Physical evidence of malnutrition found to be correlated with lower IQs in studies conducted in Africa, Mexico, and Guatemala have not been found in even the poorest and lowest IQ segments of the American Negro population. On the basis of present evidence, the hypothesis that lower average Negro IQ is due to poor nutrition is not tenable.

The nutritional and health-care status of Indian children, as indicated by much higher rates of infant mortality, is much poorer than that of Negroes; yet, Indian children in the first grade in school—age 6—have been found to score about one *SD* above Negroes on nonverbal ability tests.

8. Prenatal and Perinatal Disadvantages

The higher rate of fetal loss and infant mortality in the Negro population may indicate disadvantages related to prenatal health care of the mother and undesirable conditions attending birth. These conditions prevail in the poorer segment of the Negro population and probably contribute to the incidence of neurological handicap among Negro children. All of the causes of high fetal loss, however, are not understood, for there are some relatively disadvantaged populations which have shown lower rates of fetal loss than is found in the white majority—Orientals, for example. There is now evidence that the degree of genetic heterogeneity of the fetus' ancestors is directly related

to the probability of fetal loss, and thus genetic factors may be involved even in this seemingly environmental phenomenon. Disadvantaging forms of birth trauma such as anoxia, low birth weight and prematurity are reflected in subnormal performance on infant tests of perceptual-motor development. But large representative samples of Negro children show no depression of scores on these tests and generally perform at slightly higher levels than middle-class white children. Prenatal and perinatal factors, though differing in Negro and white populations, do not begin to account for such phenomena as the six times higher rate of mental retardation—IQs below 70—in the Negro than in the white population. Unless one hypothesizes the existence of genetic factors, in the vast majority of cases the causes of the mental retardation must be categorized as "unknown" or "unidentified."

VII. Educational Implications

At present, neither I nor anyone else, I'm afraid, has any more than rather general notions concerning the educational implications of the wide range of apparent differences in educability in our population. Since the heredity-environment issue is not likely to reach a general consensus among qualified scientists for quite some time to come and after much more genetical and psychological research has been completed, it is probably wise for educators to assume an openly agnostic position with regard to the genetic issue as it involves racial differences, at the same time recognizing that whatever may be the causes of the difference, we do not at present know of any measures or methods within the power of the schools that will appreciably or permanently diminish either individual or group differences in intelligence or scholastic achievement. There is fundamentally, in my opinion, no difference, psychologically and genetically, between individual differences and group differences. Individual differences often simply get tabulated so as to show up as group differences—between schools in different neighborhoods, between different racial groups, between cities and regions. They then become a political and ideological, not just a psychological, matter. To reduce the social tensions that arise therefrom, we see proposals to abolish aptitude and achievement testing, grading, grade placement, special classes for the educationally retarded and the academically gifted, neighborhood schools, the classroom as the instructional unit, the academic curriculum, and even our whole system of education. There may be merit in some of these proposals. But I think they are too often aimed at covering up problems rather than coming to grips with them. We can urge doing away with classification and groups, and enforce laws against racial discrimination in educational opportunities and employment and housing; we can and must insist upon considering only persons' individual characteristics rather than their group membership as a basis for educational treatment and in social relations in general. Well and good. I trust there is no disagreement on this. What we may not accomplish by these means, however, is equality of performance in school or in the acquisition of certain skills deemed valuable by society and rewarded accordingly. If we repeatedly look for the causes of differences in ability to acquire an educationally valued skill such as reading, for example, in the

external environment and are hard put to find a convincing explanation there, but we also refuse to consider any other than external factors as possible causes of these differences, perhaps we only sow the seeds of a kind of social paranoia—a need to find strictly external causes to blame for the observed differences.

To seek the answers to these questions and yet to worry about their far reaching implications: ~~surely this~~ is the scientist's moral dilemma. I don't claim to have the solution.

In terms of what we now know in educational research and in terms of what seems immediately feasible, I would suggest further consideration of three main educational approaches. They are not at all mutually exclusive. The desirability and necessity of eliminating racial discrimination and of generally improving the environmental conditions and educational and occupational opportunities of all disadvantaged persons in the population are taken for granted. These approaches have nothing to do with race *per se*, but are concerned with individual differences in those characteristics most relevant to educability. Their success in improving the benefits of education to Negro children, however, may depend in part upon recognizing that racial differences in the distribution of educationally relevant abilities are not mainly a result of discrimination and unequal environmental conditions. None of the approaches that seem to me realistic is based on the expectation of the schools' significantly changing children's basic intelligence.

1. Seeking Aptitude X Training Interactions

This means that some children may learn better by one method than by another and that the best method may be quite different for different children, depending on their particular aptitudes or other personological characteristics. It implies that the same educational goals can be accomplished to the same degree for children of different abilities provided the right instructional variations are found. This is merely a hope, and the relevant research so far gives little basis for optimism that such aptitude X training interactions will be found which can overcome to any marked degree the importance of IQ level for educability. But since this type of research has been underway only a few years, it is much too soon to discount the possibilities it may turn up—especially if one expects not miracles, but only positive, if modest, benefits from this approach.

2. Greater Attention to Learning Readiness

The concept of developmental readiness for various kinds of school learning has been too neglected in recent educational trends, which have been dominated by the unproved notion that the earlier something can be taught to a child, the better. Forced early learning, prior to some satisfactory level of readiness—which will differ markedly from one child to another—could cause learning blocks which later on practically defy remediation. The more or less uniform lock-step sequencing of educational experiences may have to be drastically modified for the benefit of many children, but the recent massive insistence on "earliness" and equality of educational treatment of all children has

militated against large-scale research on the implications of readiness for children with below-average educability within the traditional school system.

3. Greater Diversity of Curricula and Goals

Public schools which aim to serve the entire population, must move beyond narrow conceptions of scholastic achievement to find a greater diversity of ways for children over the entire range of abilities to benefit from their schooling—to benefit especially in ways that will be to their advantage when they are out of school. The academic goals of schooling are so ingrained in our thinking and our values that it will probably call for radical efforts to modify public education in ways such that it will maximally benefit large numbers of children with very limited aptitude for academic achievement. I believe that a well-intentioned but misconceived social egalitarian ideology has prevented public education in the United States from facing up to this challenge.

The belief that equality of educational opportunity should necessarily lead to equality of performance, I believe, is proving to be a false hope. It is the responsibility of scientific research in genetics, psychology, and education to determine the basis for realistic solutions to the problems of universal public education. Though it may be premature to prescribe at present, I venture the prediction that future solutions will take the form not so much of attempting to minimize differences in scholastic aptitudes and motivation, but of creating a greater diversity of curricula, instructional methods, and educational goals and values that will make it possible for children—ranging over a wider spectrum of abilities and proclivities—to genuinely benefit from their years in school. The current zeitgeist of environmentalist equalitarianism has all but completely stifled our thinking along these lines.

And I believe the magnitude and urgency of the problem are such as to call for quite radical thinking if the educational system is truly to serve the whole of society. We have invested so much for so long in trying to equalize scholastic performance that we have given little or no thought to finding ways of diversifying schools to make them rewarding to everyone while not attempting to equalize everyone's performance in a common curriculum. Recommendations have almost always taken the form of asking what next we might try to make children who in the present school system do not flourish academically become more like those who do. The emphasis has been more on changing children than on revamping the system. A philosophy of equalization, however laudable its ideals, cannot work if it is based on false premises, and no amount of propaganda can make it appear to work. Its failures will be forced upon everyone. Educational pluralism of some sort, encompassing a variety of very different educational curricula and goals, I think, will be the inevitable outcome of the growing realization that the schools are not going to eliminate human differences. Rather than making over a large segment of the school population so they will not be doomed to failure in a largely antiquated, elitist-oriented, educational system—which originally evolved to serve

only a relatively small segment of society—the educational system will have to be revamped in order to benefit everyone who is required by the society to attend school. It seems incredible that a system can still survive which virtually guarantees frustration and failure for a large proportion of the children it should intend to serve. From all the indications, public education in such a form will not much longer survive.

But we should not fail to recognize that to propose radical diversity in accord with individual differences in abilities and interests, as contrasted with uniformity of educational treatment, puts society between Scylla and Charybdis in terms of insuring for all individuals equality of opportunity for the diversity of educational paths. The surest way to maximize the benefits of schooling to all individuals and at the same time to make the most of a society's human resources is to insure equality of educational opportunity for all its members. Monolithic educational goals and uniformity of approaches guarantee unnecessary frustration and defeat for many. On the other hand, educational pluralism runs the risk that social, economic, ethnic background or geographic origin, rather than each child's own characteristic, might determine the educational paths available to him. The individual characteristics appropriate for any one of a variety of educational paths and goals are to be found everywhere, in every social stratum, ethnic group, and neighborhood. Academic aptitudes and special talents should be cultivated wherever they are found, and a wise society will take all possible measures to insure this to the greatest possible extent. At the same time, those who are poor in the tradition academic aptitudes cannot be left by the wayside. Suitable means and goals must be found for making their years of schooling rewarding to them, if not in the usual academic sense, then in ways that can better their chances for socially useful and self-fulfilling roles as adults.

How Much Can We Boost IQ and Scholastic Achievement?

ARTHUR R. JENSEN

University of California, Berkeley

How Much Can We Boost IQ and Scholastic Achievement?

Arthur R. Jensen

University of California, Berkeley

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

How Much Can We Boost IQ and Scholastic Achievement?

ARTHUR R. JENSEN

University of California, Berkeley

Arthur Jensen argues that the failure of recent compensatory education efforts to produce lasting effects on children's IQ and achievement suggests that the premises on which these efforts have been based should be reexamined.

He begins by questioning a central notion upon which these and other educational programs have recently been based: that IQ differences are almost entirely a result of environmental differences and the cultural bias of IQ tests. After tracing the history of IQ tests, Jensen carefully defines the concept of IQ, pointing out that it appears as a common factor in all tests that have been devised thus far to tap higher mental processes.

Having defined the concept of intelligence and related it to other forms of mental ability, Jensen employs an analysis of variance model to explain how IQ can be separated into genetic and environmental components. He then discusses the concept of "heritability," a statistical tool for assessing the degree to which individual differences in a trait like intelligence can be accounted for by genetic factors. He analyzes several lines of evidence which suggest that the heritability of intelligence is quite high (i.e., genetic factors are much more important than environmental factors in producing IQ differences).

After arguing that environmental factors are not nearly as important in determining IQ as are genetic factors, Jensen proceeds to analyze the environmental influences which may be most critical in determining IQ. He concludes

Harvard Educational Review Vol. 39 No. 1 Winter 1969

that prenatal influences may well contribute the largest environmental influence on IQ. He then discusses evidence which suggests that social class and racial variations in intelligence cannot be accounted for by differences in environment but must be attributed partially to genetic differences.

After he has discussed the influence of the distribution of IQ in a society on its functioning, Jensen examines in detail the results of educational programs for young children, and finds that the changes in IQ produced by these programs are generally small. A basic conclusion of Jensen's discussion of the influence of environment on IQ is that environment acts as a "threshold variable." Extreme environmental deprivation can keep the child from performing up to his genetic potential, but an enriched educational program cannot push the child above that potential.

Finally, Jensen examines other mental abilities that might be capitalized on in an educational program, discussing recent findings on diverse patterns of mental abilities between ethnic groups and his own studies of associative learning abilities that are independent of social class. He concludes that educational attempts to boost IQ have been misdirected and that the educational process should focus on teaching much more specific skills. He argues that this will be accomplished most effectively if educational methods are developed which are based on other mental abilities besides I.Q.

Because of the controversial nature of Dr. Jensen's article, the Spring Issue of the Review *will feature a discussion of the article by five psychologists: Carl Bereiter, Lee Cronbach, James Crow, David Elkind, and J. McVicker Hunt. Readers are also invited to react.*

The Failure of Compensatory Education

Compensatory education has been tried and it apparently has failed.

Compensatory education has been practiced on a massive scale for several years in many cities across the nation. It began with auspicious enthusiasm and high hopes of educators. It had unprecedented support from Federal funds. It had theoretical sanction from social scientists espousing the major underpinning of its rationale: the "deprivation hypothesis," according to which academic lag is mainly the result of social, economic, and educational deprivation and discrimination—an hypothesis that has met with wide, uncritical acceptance in the atmosphere of society's growing concern about the plight of minority groups and the economically disadvantaged.

The chief goal of compensatory education—to remedy the educational lag of disadvantaged children and thereby narrow the achievement gap between "minority" and "majority" pupils—has been utterly unrealized in any of the large compensatory education programs that have been evaluated so far. On the basis of a nationwide survey and evaluation of compensatory education programs, the United States Commission on Civil Rights (1967) came to the following conclusion:

The Commission's analysis does not suggest that compensatory education is incapable of remedying the effects of poverty on the academic achievement of individual children. There is little question that school programs involving expenditures for cultural enrichment, better teaching, and other needed educational services can be helpful to disadvantaged children. The fact remains, however, that none of the programs appear to have raised significantly the achievement of participating pupils, as a group, within the period evaluated by the Commission. (p. 138)

The Commission's review gave special attention to compensatory education in majority-Negro schools whose programs "were among the most prominent and included some that have served as models for others." The Commission states: "A principal objective of each was to raise the academic achievement of disadvantaged children. Judged by this standard the programs did not show evidence of much success" (p. 138).[1]

Why has there been such uniform failure of compensatory programs wherever they have been tried? What has gone wrong? In other fields, when bridges do not stand, when aircraft do not fly, when machines do not work, when treatments do not cure, despite all conscientious efforts on the part of many persons to make them do so, one begins to question the basic assumptions, principles, theories, and hypotheses that guide one's efforts. Is it time to follow suit in education?

[1] Some of the largest and most highly publicized programs of compensatory education that have been held up as models but which produced absolutely no significant improvement in the scholastic achievement of disadvantaged students are: the *Banneker Project* in St. Louis (8 years), *Higher Horizons* in New York (5 years), More Effective Schools in New York (3 years), and large-scale programs in Syracuse, Seattle, Philadelphia, Berkeley, and a score of other cities (for detailed reports see U.S. Commission on Civil Rights, 1967, pp. 115-140).

Reports on Project Head Start indicate that initial gains of 5 to 10 points in IQ on conventional intelligence tests are a common finding, but this gain usually does not hold up through the first year of regular schooling. More positive claims for the efficacy of Head Start involve evidence of the detection and correction of medical disabilities in disadvantaged preschool children and the reportedly favorable effects of the program on children's self-confidence, motivation, and attitudes toward school.

The theory that has guided most of these compensatory education programs, sometimes explicitly, sometimes implicitly, has two main complementary facets: one might be called the "average children concept," the other the "social deprivation hypothesis."

The "average children" concept is essentially the belief that all children, except for a rare few born with severe neurological defects, are basically very much alike in their mental development and capabilities, and that their apparent differences in these characteristics as manifested in school are due to rather superficial differences in children's upbringing at home, their preschool and out-of-school experiences, motivations and interests, and the educational influences of their family background. All children are viewed as basically more or less homogeneous, but are seen to differ in school performance because when they are out of school they learn or fail to learn certain things that may either help them or hinder them in their school work. If all children could be treated more alike early enough, long before they come to school, then they could all learn from the teacher's instruction at about the same pace and would all achieve at much the same level, presumably at the "average" or above on the usual grade norms.

The "social deprivation hypothesis" is the allied belief that those children of ethnic minorities and the economically poor who achieve "below average" in school do so mainly because they begin school lacking certain crucial experiences which are prerequisites for school learning—perceptual, attentional, and verbal skills, as well as the self-confidence, self-direction, and teacher-oriented attitudes conducive to achievement in the classroom. And they lack the parental help and encouragement needed to promote academic achievement throughout their schooling. The chief aim of preschool and compensatory programs, therefore, is to make up for these environmental lacks as quickly and intensively as possible by providing the assumedly appropriate experiences, cultural enrichment, and training in basic skills of the kind presumably possessed by middle-class "majority" children of the same age.

The success of the effort is usually assessed in one or both of two ways: by gains in IQ and in scholastic achievement. The common emphasis on gains in IQ is probably attributable to the fact that it can be more efficiently "measured" than scholastic achievement, especially if there is no specific "achievement" to begin with. The IQ test can be used at the very beginning of Headstart, kindergarten, or first grade as a "pre-test" against which to assess "post-test" gains. IQ gains, if they occur at all, usually occur rapidly, while achievement is a long-term affair. And probably most important, the IQ is commonly interpreted as indicative of

a more general kind of intellectual ability than is reflected by the acquisition of specific scholastic knowledge and skills. Since the IQ is known to predict scholastic performance better than any other single measurable attribute of the child, it is believed, whether rightly or wrongly, that if the child's IQ can be appreciably raised, academic achievement by and large will take care of itself, given normal motivation and standard instruction. Children with average or above-average IQs generally do well in school without much special attention. So the remedy deemed logical for children who would do poorly in school is to boost their IQs up to where they can perform like the majority—in short to make them all at least "average children." Stated so bluntly, the remedy may sound rather grim, but this is in fact essentially what we are attempting in our special programs of preschool enrichment and compensatory education. This simple theme, with only slight embellishments, can be found repeated over and over again in the vast recent literature on the psychology and education of children called culturally disadvantaged.

So here is where our diagnosis should begin—with the concept of the IQ: how it came to be what it is; what it "really" is; what makes it vary from one individual to another; what can change it, and by what amount.

The Nature of Intelligence

The nature of intelligence is one of the vast topics in psychology. It would be quite impossible to attempt to review here the main theoretical issues and currents of thought in this field. Large volumes have been written on the subject (e.g., Guilford, 1967; Stoddard, 1943), to say nothing of the countless articles. An enlightening brief account of the history of the concept of intelligence has been presented by Sir Cyril Burt (1968). The term "intelligence," as used by psychologists, is itself of fairly recent origin. Having been introduced as a technical term in psychology near the turn of the century, it has since filtered down into common parlance, and therefore some restriction and clarification of the term as it will be used in the following discussion is called for.

Disagreements and arguments can perhaps be forestalled if we take an operational stance. First of all, this means that probably the most important fact about intelligence is that we can measure it. Intelligence, like electricity, is easier to measure than to define. And if the measurements bear some systematic relationships to other data, it means we can make meaningful statements about the phenomenon we are measuring. There is no point in arguing the question to which

there is no answer, the question of what intelligence *really* is. The best we can do is to obtain measurements of certain kinds of behavior and look at their relationships to other phenomena and see if these relationships make any kind of sense and order. It is from these orderly relationships that we can gain some understanding of the phenomena.

But how did the instruments by which we measure intelligence come about in the first place? The first really useful test of intelligence and the progenitor of nearly all present-day intelligence tests was the Metrical Scale of Intelligence devised in 1905 by Binet and Simon. A fact of great but often unrealized implications is that the Binet-Simon test was commissioned by the Minister of Public Instruction in Paris for the explicit purpose of identifying children who were likely to fail in school. It was decided they should be placed in special schools or classes before losing too much ground or receiving too much discouragement. To the credit of Binet and Simon, the test served this purpose quite well, and it is now regarded as one of the major "breakthroughs" in the history of psychology. Numerous earlier attempts to devise intelligence tests were much less successful from a practical standpoint, mainly because the kinds of functions tested were decided upon in terms of early theoretical notions about the basic elements of "mind" and the "brass instrument" laboratory techniques for measuring these elemental functions of consciousness, which were then thought to consist of the capacity for making fine sensory discriminations in the various sensory modalities. Although these measurements were sufficiently reliable, they bore little relationship to any "real life" or "common sense" criteria of behavior ranging along a "dull"—"bright" continuum. The psychological sagacity of Binet and Simon as test constructors derived largely from their intimate knowledge and observation of the behavior of young children and of what, precisely, teachers expected of them in school. Binet and Simon noted the characteristics distinguishing those children described by their teachers as "bright" from those described as "dull," and, from these observations and considerable trial-and-error, they were finally able to make up a graded series of test items that not only agreed with teachers' judgments of children's scholastic capabilities but could make the discriminations more finely and more accurately than any single teacher could do without prolonged observation of the child in class. The Binet-Simon scale has since undergone many revisions and improvements, and today, in the form developed by Terman, known as the Stanford-Binet Intelligence Scale, it is generally regarded as the standard for the measurement of intelligence.

But the important point I wish to emphasize here is that these Binet tests, and

77-942 O - 72 - 6

in effect all their descendants, had their origin in the educational setting of the Paris schools of 1900, and the various modifications and refinements they have undergone since then have been implicitly shaped by the educational traditions of Europe and North America. The content and methods of instruction represented in this tradition, it should be remembered, are a rather narrow and select sample of all the various forms of human learning and of the ways of imparting knowledge and skills. The instructional methods of the traditional classroom were not invented all in one stroke, but evolved within an upper-class segment of the European population, and thus were naturally shaped by the capacities, culture, and needs of those children whom the schools were primarily intended to serve. At least implicit in the system as it originally developed was the expectation that not all children would succeed. These methods of schooling have remained essentially unchanged for many generations. We have accepted traditional instruction so completely that it is extremely difficult even to imagine, much less to put into practice, any radically different forms that the education of children could take. Our thinking almost always takes as granted such features as beginning formal instruction at the same age for all children (universally between ages five and six), instruction of children in groups, keeping the same groups together in lock step fashion through the first several years of schooling, and an active-passive, showing-seeing, telling-listening relationship between teacher and pupils. Satisfactory learning occurs under these conditions only when children come to school with certain prerequisite abilities and skills: an attention span long enough to encompass the teacher's utterances and demonstrations, the ability voluntarily to focus one's attention where it is called for, the ability to comprehend verbal utterances and to grasp relationships between things and their symbolic representations, the ability to inhibit large-muscle activity and engage in covert "mental" activity, to repeat instruction to oneself, to persist in a task until a self-determined standard is attained—in short, the ability to engage in what might be called self-instructional activities, without which group instruction alone remains ineffectual.

The interesting fact is that, despite all the criticisms that can easily be leveled at the educational system, the traditional forms of instruction have actually worked quite well for the majority of children. And the tests that were specifically devised to distinguish those children least apt to succeed in this system have also proved to do their job quite well. The Stanford-Binet and similar intelligence tests predict various measures of scholastic achievement with an average validity coefficient of about .5 to .6, and in longitudinal data comprising intelligence

test and achievement measures on the same children over a number of years, the multiple correlation between intelligence and scholastic achievement is almost as high as the reliability of the measures will permit.

The Generality and Limitations of Intelligence

If the content and instructional techniques of education had been markedly different from what they were in the beginning and, for the most part, continue to be, it is very likely that the instruments we call intelligence tests would also have assumed a quite different character. They might have developed in such a way as to measure a quite different constellation of abilities, and our conception of the nature of intelligence, assuming we still called it by that name, would be correspondingly different. This is why I think it so important to draw attention to the origins of intelligence testing.

But in granting that the measurement and operational definitions of intelligence had their origins in a school setting and were intended primarily for scholastic purposes, one should not assume that intelligence tests measure *only* school learning or cultural advantages making for scholastic success and fail to tap anything of fundamental psychological importance. The notion is sometimes expressed that psychologists have mis-aimed with their intelligence tests. Although the tests may predict scholastic performance, it is said, they do not *really* measure intelligence—as if somehow the "real thing" has eluded measurement and perhaps always will. But this is a misconception. We *can* measure intelligence. As the late Professor Edwin G. Boring pointed out, intelligence, by definition, is what intelligence tests measure. The trouble comes only when we attribute more to "intelligence" and to our measurements of it than do the psychologists who use the concept in its proper sense.

The idea of intelligence has justifiably grown considerably beyond its scholastic connotations. Techniques of measurement not at all resembling the tasks of the Binet scale and in no way devised with the idea of predicting scholastic performance can also measure approximately the same intelligence as measured by the Binet scale. The English psychologist Spearman devoted most of his distinguished career to studying the important finding that almost any and every test involving any kind of complex mental activity correlates positively and substantially with any and every other test involving complex mental activity, regardless of the specific content or sensory modality of the test. Spearman noted that if the tests called for the operation of "higher mental processes," as opposed to sheer sensory acuity, reflex behavior, or the execution of established habits, they showed

positive intercorrelations, although the tests bore no superficial resemblance to one another. They might consist of abstract figures involving various spatial relationships, or numerical problems, or vocabulary, or verbal analogies. For example, a vocabulary test shows correlations in the range of .50 to .60 with a test that consists of copying sets of designs with colored blocks; and a test of general information correlates about .50 with a test that involves wending through a printed maze with a pencil. Countless examples of such positive correlations between seemingly quite different tests can be found in the literature on psychological tests. Spearman made them the main object of his study. To account for the intercorrelations of "mental" tests, he hypothesized the existence of a single factor common to all tests involving complex mental processes. All such tests measure this common factor to some degree, which accounts for the intercorrelations among all the tests. Spearman called the common factor "general intelligence" or simply *g*. And he invented the method known as factor analysis to determine the amount of *g* in any particular test. He and his students later developed tests, like Raven's Progressive Matrices and Cattell's Culture Fair Tests of g, which measure *g* in nearly pure form. We should not reify *g* as an entity, of course, since it is only a hypothetical construct intended to explain covariation among tests. It is a hypothetical source of variance (individual differences) in test scores. It can be regarded as the nuclear operational definition of intelligence, and when the term intelligence is used it should refer to *g*, the factor common to all tests of complex problem solving.

In examining those tests most heavily loaded with *g*, Spearman characterized the mental processes which they seemed to involve as "the ability to educe relations and correlates"—that is, to be able to see the general from the particular and the particular as an instance of the general. A similar definition of intelligence was expressed by Aquinas, as "the ability to combine and separate"—to see the difference between things which seem similar and to see the similarities between things which seem different.These are essentially the processes of abstraction and conceptualization. Tasks which call for problem solving requiring these processes are usually the best measures of *g*. Despite numerous theoretical attacks on Spearman's basic notion of a general factor, *g* has stood like a rock of Gibralter in psychometrics, defying any attempt to construct a test of complex problem solving which excludes it.

Standard intelligence scales such as the Binet and the Wechsler are composed of a dozen or so subtests which differ obviously in their superficial appearance: vocabulary, general information, memory span for digits, block designs, figure

copying, mazes, form boards, and so on. When the intercorrelations among a dozen or more such tests are subjected to a factor analysis or principal components analysis, some 50 percent or more of the total individual differences variance in all the tests is usually found to be attributable to a general factor common to all the tests. Thus, when we speak of intelligence it is this general factor, rather than any single test, that we should keep in mind.

Attempts to assess age differences in intelligence or mental development which rely on complex techniques that bear little formal resemblance to the usual intelligence tests still manage to measure *g* more than anything else. Piaget's techniques for studying mental growth, for example, are based largely on the child's development of the concepts of invariance and conservation of certain properties —number, area, and volume. When a large variety of Piaget tasks are factor analyzed along with standard psychometric tests, including the Stanford-Binet and Raven's Progressive Matrices, is it found that the Piaget tasks are loaded on the general factor to about the same extent as the psychometric tests (Vernon, 1965). That is to say, children fall into much the same rank order of ability on all these cognitive tests. Tuddenham (1968) has developed a psychometric scale of intelligence based entirely upon Piaget's theory of cognitive development. The test makes use of ten of the techniques developed by Piaget for studying conservation, seriation, reversal of perspective, and so on. Performance on these tasks shows about the same relationship to social class and race differences as is generally found with the Stanford-Binet and Wechsler scales. It seems evident that what we call general intelligence can be manifested in many different forms and thus permits measurement by a wide variety of techniques. The common feature of all such intercorrelated tests seems to be their requirement of some form of "reasoning" on the part of the subject—some active, but usually covert, transformation or manipulation of the "input" (the problem) in order to arrive at the "output" (the answer).

The conceptually most pure and simple instance of this key aspect of intelligence is displayed in the phenomenon known as cross-modal transfer. This occurs when a person to whom some particular stimulus is exposed in one sensory modality can then recognize the same stimulus (or its essential features) in a different sensory modality. For example, show a person a number of differently shaped wooden blocks, then point to one, blindfold the person, shuffle the blocks, and let the person find the indicated block by using his sense of touch. Or "write" in bold strokes any letter of the alphabet between a child's shoulder blades. It will be a completely unique stimulus input for the child, never encountered before and

never directly conditioned to any verbal response. Yet, most children, provided they already know the alphabet, will be able to name the letter. There are no direct neural connections between the visual and the tactile impressions of the stimulus, and, although the child's naming of the letter has been conditioned to the visual stimulus, the tactile stimulus has been associated with neither the visual stimulus nor the verbal response. How does the child manage to show the cross modal transfer? Some central symbolic or "cognitive" processing mechanism is involved, which can abstract and compare properties of "new" experiences with "old" experiences and thereby invest the "new" with meaning and relevance. Intelligence is essentially characterized by this process.

Is g Unitary or Divisible?

It is only when the concept of *g* is attributed meaning above and beyond that derived from the factor analytic procedures from which it gains its strict technical meaning that we run into the needless argument over whether *g* is a unitary ability or a conglomerate of many subabilities, each of which could be measured independently. We should think of *g* as a "source" of individual differences in scores which is common to a number of different tests. As the tests change, the nature of *g* will also change, and a test which is loaded, say, .50 on *g* when factor analyzed among one set of tests may have a loading of .20 or .80, or some other value, when factor analyzed among other sets of tests. Also, a test which, in one factor analysis, measures only *g* and nothing else, may show that it measures *g* and one or more other factors when factor analyzed in connection with a new set of tests. In other words, *g* gains its meaning from the tests which have it in common. Furthermore, no matter how simple or "unitary" a test may appear to be, it is almost always possible to further fractionate the individual differences variance into smaller subfactors. I have been doing this in my laboratory with respect to a very simple and seemingly "unitary" ability, namely, digit span (Jensen, 1967b). Changing the rate of digit presentation changes the rank order of subjects in their ability to recall the digits. So, too, does interposing a 10-second delay between presentation and recall, and interpolating various distractions ("retroactive inhibition") between presentation and recall, and many other procedural variations of the digit span paradigm. Many—but, significantly, not all—of these kinds of manipulations introduce new dimensions or factors of individual differences. It is likely that when we finally get down to the irreducible "atoms" of memory span ability, so to speak, if we ever do get there, the elements that make up mem-

ory span ability will not themselves even resemble what we think of as abilities in the usual sense of the term. And so probably the same would be true not only for digit span, but for any of the subtests or items that make up intelligence tests.

A simple analogy in the physical realm may help to make this clear. If we are interested in measuring general athletic ability, we can devise a test consisting of running, ball throwing, batting, jumping, weight lifting, and so on. We can obtain a "score" on each one of these and the total for any individual is his "general athletic ability" score. This score would correspond to the general intelligence score yielded by tests like the Stanford-Binet and the Wechsler scales.

Or we can go a step further in the refinement of our test procedure and intercorrelate the scores on all these physical tasks, factor analyze the intercorrelations, and examine the general factor, if indeed there is one. Assuming there is, we would call it "general athletic ability." It would mean that on all of the tasks, persons who excelled on one also tended to be superior on the others. And we would note that some tasks were more "loaded" with this general factor than others. We could then weight the subtest scores in proportion to their loading on *g* and then add them up. The total, in effect, is a "factor score," and gives us a somewhat more justifiable measure of "general athletic ability," since it represents the one source of variation that all the athletic skills in our test battery share in common.

To go still further, let us imagine that the running test has the highest loading on *g* in this analysis. To make the issue clear-cut, let us say that all its variance is attributable to the *g* factor. Does this mean that running ability is not further analyzable into other components? *No, it simply means that the components into which running can be analyzed are not separately or independently manifested in either the running test or the other tests in the battery.* But we can measure these components of running ability independently, if we wish to: total leg length, the ratio of upper to lower leg length, strength of leg muscles, physical endurance, "wind" or vital capacity, ratio of body height to weight, degree of mesomorphic body build, specific skills such as starting speed—all are positively correlated with running speed. And if we intercorrelate these measures and factor analyze the correlations, we would probably find a substantial general factor common to all these physical attributes, name it what you will. We could combine the measures on these various physical traits into a weighted composite score which would predict running ability as measured by the time the person takes to cross the finish line. The situation seems very similar to the analysis of the psychological processes that make up "general intelligence."

Fluid and Crystallized Intelligence

Raymond B. Cattell (1963) has made a conceptually valid distinction between two aspects of intelligence, *fluid* and *crystallized*. Standard intelligence tests generally measure both the fluid and crystallized components of *g*, and, since the two are usually highly correlated in a population whose members to a large extent share a common background of experience, culture, and education, the fluid and crystallized components may not always be clearly discernible as distinct factors. Conceptually, however, the distinction is useful and can be supported empirically under certain conditions. *Fluid* intelligence is the capacity for new conceptual learning and problem solving, a general "brightness" and adaptability, relatively independent of education and experience, which can be invested in the particular opportunities for learning encountered by the individual in accord with his motivations and interests. Tests that measure mostly fluid intelligence are those that minimize cultural and scholastic content. Cattell's Culture Fair Tests and Raven's Progressive Matrices are good examples. *Crystallized* intelligence, in contrast, is a precipitate out of experience, consisting of acquired knowledge and developed intellectual skills. Fluid and crystallized intelligence are naturally correlated in a population sharing a common culture, because the acquisition of knowledge and skills in the first place depends upon fluid intelligence. While fluid intelligence attains its maximum level in the late teens and may even begin to decline gradually shortly thereafter, crystallized intelligence continues to increase gradually with the individual's learning and experience all the way up to old age.

Occupational Correlates of Intelligence

Intelligence, as we are using the term, has relevance considerably beyond the scholastic setting. This is so partly because there is an intimate relationship between a society's occupational structure and its educational system. Whether we like it or not, the educational system is one of society's most powerful mechanisms for sorting out children to assume different roles in the occupational hierarchy.

The evidence for a hierarchy of occupational prestige and desirability is unambiguous. Let us consider three sets of numbers.[8] First, the Barr scale of occupations, devised in the early 1920s, provides one set of data. Lists of 120 representative occupations, each definitely and concretely described, were given to 30 psychological judges who were asked to rate the occupations on a scale from 0 to 100

[8] I am indebted to Professor Otis Dudley Duncan (1968, pp. 80-100) for providing this information.

according to the grade of intelligence each occupation was believed to require for ordinary success. Second, in 1964, the National Opinion Research Center (NORC), by taking a large public opinion poll, obtained ratings of the *prestige* of a great number of occupations; these prestige ratings represent the average standing of each occupation relative to all the others in the eyes of the general public. Third, a rating of socioeconomic status (SES) is provided by the *1960 Census of Population: Classified Index of Occupations and Industries*, which assigns to each of the hundreds of listed occupations a score ranging from 0 to 96 as a composite index of the average income and educational level prevailing in the occupation.

The interesting point is the set of correlations among these three independently derived occupational ratings.

The Barr scale and the NORC ratings are correlated .91.

The Barr scale and the SES index are correlated .81.

The NORC ratings and the SES index are correlated .90.

In other words, psychologists' concept of the "intelligence demands" of an occupation (Barr scale) is very much like the general public's concept of the prestige or "social standing" of an occupation (NORC ratings), and both are closely related to an independent measure of the educational and economic status of the persons pursuing an occupation (SES index). As O. D. Duncan (1968, pp. 90-91) concludes, "... 'intelligence' is a socially defined quality and this social definition is not essentially different from that of achievement or status in the occupational sphere.... When psychologists came to propose operational counterparts to the notion of intelligence, or to devise measures thereof, they wittingly or unwittingly looked for indicators of capability to function in the system of key roles in the society." Duncan goes on to note, "Our argument tends to imply that a correlation between IQ and occupational achievement was more or less built into IQ tests, by virtue of the psychologists' implicit acceptance of the social standards of the general populace. Had the first IQ tests been devised in a hunting culture, 'general intelligence' might well have turned out to involve visual acuity and running speed, rather than vocabulary and symbol manipulation. As it was, the concept of intelligence arose in a society where high status accrued to occupations involving the latter in large measure, so that what we now *mean* by intelligence is something like the probability of acceptable performance (given the opportunity) in occupations varying in social status."

So we see that the prestige hierarchy of occupations is a reliable objective reality in our society. To this should be added the fact that there is undoubtedly some relationship between the levels of the hierarchy and the occupations' in-

trinsic interest, desirability, or gratification to the individuals engaged in them. Even if all occupations paid alike and received equal respect and acclaim, some occupations would still be viewed as more desirable than others, which would make for competition, selection, and, again, a kind of prestige hierarchy. Most persons would agree that painting pictures is more satisfying than painting barns, and conducting a symphony orchestra is more exciting than directing traffic. We have to face it: the assortment of persons into occupational roles simply is not "fair" in any absolute sense. The best we can ever hope for is that true merit, given equality of opportunity, act as the basis for the natural assorting process.

Correlation Between Intelligence and Occupational Achievement.

Because intelligence is only one of a number of qualities making for merit in any given occupation, and since most occupations will tolerate a considerable range of abilities and criteria of passable performance, it would be surprising to find a very high correlation between occupational level and IQ. Although the rank order of the *mean* IQs of occupational groups is about as highly correlated with the occupations' standing on the three "prestige" ratings mentioned above as the ratings are correlated among themselves, there is a considerable dispersion of IQs *within* occupations. The IQ spread increases as one moves down the scale from more to less skilled occupations (Tyler, 1965, pp. 338-339). Thus, the correlation, for example, between scores on the Army General Classification Test, a kind of general intelligence test, and status ratings of the civilian occupations of 18,782 white enlisted men in World War II was only .42. Since these were mostly young men, many of whom had not yet completed their education or established their career lines, the correlation of .42 is lower than one would expect in the civilian population. Data obtained by the U.S. Employment Service in a civilian population shows a correlation of .55 between intelligence and occupational status, a value which, not surprisingly, is close to the average correlation between intelligence and scholastic achievement (Duncan, et al., 1968, pp. 98-101). Although these figures are based on the largest samples reported in the literature and are therefore probably the most reliable statistics, they are not as high as the correlations found in some other studies. Two studies found, for example, that IQs of school boys correlated .57 and .71 with their occupational status 14 and 19 years later, respectively (Tyler, 1965, p. 343). It is noteworthy that the longer interval showed the higher correlation.

Duncan's (1968) detailed analysis of the nature of the relationship between intelligence and occupational status led him to the conclusion that "the bulk of the

influence of intelligence on occupation is indirect, via education." If the correlation of intelligence with education and of education with occupation is, in effect, "partialled out," the remaining "direct" correlation between intelligence and occupation is almost negligible. But Duncan points out that this same type of analysis (technically known as "path coefficients analysis") also reveals the interesting and significant finding that intelligence plays a relatively important part as a cause of differential *earnings*. Duncan concludes: ". . . men with the same schooling and in the same line of work are differentially rewarded in terms of mental ability" (1968, p. 118).

Correlations Between Intelligence and Job Performance Within Occupations

Intelligence, via education, has its greatest effect in the assorting of individuals into occupational roles. Once they are in those roles, the importance of intelligence per se is less marked. Ghiselli (1955) found that intelligence tests correlate on the average in the range of .20 to .25 with ratings of actual proficiency on the job. The speed and ease of training for various occupational skills, however, show correlations with intelligence averaging about .50, which is four to five times the predictive power that the same tests have in relation to work proficiency *after* training. This means that, once the training hurdle has been surmounted, many factors besides intelligence are largely involved in success on the job. This is an important fact to keep in mind at later points in this article.

Is Intelligence "Fixed"?

Since the publication of J. McV. Hunt's well-known and influential book. *Intelligence and Experience* (1961), the notion of "fixed intelligence" has assumed the status of a popular cliché among many speakers and writers on intelligence, mental retardation, cultural disadvantage, and the like, who state, often with an evident sense of virtue and relief, that modern psychology has overthrown the "belief in fixed intelligence." This particular bugaboo seems to have loomed up largely in the imaginations of those who find such great satisfaction in the idea that "fixed intelligence" has been demolished once and for all.

Actually, there has been nothing much to demolish. When we look behind the rather misleading term "fixed intelligence," what we find are principally two real and separate issues, each calling for empirical study rather than moral philosophizing. Both issues lend themselves to empirical investigation and have long been subjects of intensive study. The first issue concerns the genetic basis of individual differences in intelligence; the second concerns the stability or constancy of the IQ throughout the individual's lifetime.

Genotype and Phenotype. Geneticists have avoided confusion and polemics about the issue of whether or not a given trait is "fixed" by asking the right question in the first place: how much of the variation (i.e., individual differences) in a particular trait or characteristic that we observe or measure (i.e., the *phenotype*) in a given population can we account for in terms of variation in the genetic factors (i.e., the *genotype*) affecting the development of the characteristic?

The genetic factors are completely laid down when the parental sperm and ovum unite. Thus the individual's genotype, by definition, is "fixed" at the moment of conception. Of course, different potentials of the genotype may be expressed at different times in the course of the individual's development. But beyond conception, whatever we observe or measure of the organism is a phenotype, and this, by definition, is *not* "fixed." The phenotype is a result of the organism's internal genetic mechanisms established at conception and all the physical and social influences that impinge on the organism throughout the course of its development. Intelligence is a phenotype, not a genotype, so the argument about whether or not intelligence is "fixed" is seen to be spurious.

The really interesting and important question, which can be empirically answered by the methods of quantitative genetics, is: what is the correlation between genotypes and phenotypes at any given point in development? For continuous or metrical characteristics such as height and intelligence, the correlation, of course, can assume any value between 0 and 1. The square of the correlation between genotype and phenotype is technically known as the *heritability* of the characteristic, a concept which is discussed more fully in a later section.

The Stability of Intelligence Measures. The second aspect of the issue of "fixed intelligence" concerns the stability of intelligence measurements throughout the course of the individual's development. Since intelligence test scores are not points on an absolute scale of measurement like height and weight, but only indicate the individual's relative standing with reference to a normative population, the question we must ask is: To what extent do individuals maintain their standing relative to one another in measured intelligence over the course of time? The answer is to be found in the correlation between intelligence test scores on a group of persons at two points in time. Bloom (1964) has reviewed the major studies of this question and the evidence shows considerable consistency.

In surveying all the correlations reported in the literature between intelligence measured on the same individuals at two points in time, I have worked out a

simple formula that gives a "best fit" to all these data. The formula has the virtue of a simple mnemonic, being much easier to remember than all the tables of correlations reported in the literature and yet being capable of reproducing the correlations with a fair degree of accuracy.

$$\hat{r}_{12} = r_{tt}\sqrt{\frac{CA_1}{CA_2}} \tag{1}$$

where $\hat{r}_{12}$ = the estimated correlation between tests given at times 1 and 2.
r_{tt} = the equivalent-forms or immediate test-retest reliability of the test.
CA_1 = the subject's chronological age at the time of the first test.
CA_2 = the subject's chronological age at the time of the second test.

Limitation: The formula holds only up to the point where CA_2 is age 10, at which time the empirical value of r_{12} approaches an asymptote, showing no appreciable increase thereafter. Beyond age 10, regardless of the interval between tests, the obtained test-retest correlations fall in the range between the test's reliability and the square of the reliability (i.e., $r_{tt} > r_{12} > r^2_{tt}$). These simple generalizations are intended simply as a means of summarizing the mass of empirical findings. They accord with Bloom's conclusion, based on his thorough survey of the published evidence, that beyond age 8, correlations between repeated tests of general intelligence, corrected for unreliability of measurement, are between +.90 and unity (Bloom, 1964, p. 61).

What these findings mean is that the IQ is not constant, but, like all other developmental characteristics, is quite variable early in life and becomes increasingly stable throughout childhood. By age 4 or 5, the IQ correlates about .70 with IQ at age 17, which means that approximately half (i.e., the square of the correlation) of the variance in adult intelligence can be predicted as early as age 4 or 5. This fact that half the variance in adult intelligence can be accounted for by age 4 has led to the amazing and widespread, but unwarranted and fallacious, conclusion that persons develop 50 percent of their mature intelligence by age 4! This conclusion, of course, does not at all logically follow from just knowing the magnitude of the correlation. The correlation between *height* at age 4 and at age 17 is also about .70, but who would claim that the square of the correlation indicated the proportion of adult height attained by age 4? The absurdity of this non sequitur is displayed in the prediction it yields: the average 4 year old boy should grow up to be 6 ft. 7 in. tall by age 17!

Intelligence has about the same degree of stability as other developmental characteristics. For example, up to age 5 or 6, height is somewhat more stable than

intelligence, and thereafter the developmental rates of height and intelligence are about equally stable, except for a period of 3 or 4 years immediately after the onset of puberty, during which height is markedly less stable than intelligence. Intelligence is somewhat more stable than total body weight over the age range from 2 to 18 years. Intelligence has a considerably more stable growth rate than measures of physical strength (Bloom, 1964, pp. 46-47). Thus, although the IQ is certainly not "constant," it seems safe to say that under normal environmental conditions it is at least as stable as developmental characteristics of a strictly physical nature.

Intelligence as a Component of Mental Ability

The term "intelligence" should be reserved for the rather specific meaning I have assigned to it, namely, the general factor common to standard tests of intelligence. Any one verbal definition of this factor is really inadequate, but, if we must define it in so many words, it is probably best thought of as a capacity for abstract reasoning and problem solving.

What I want to emphasize most, however, is that *intelligence* should not be regarded as completely synonymous with what I shall call *mental ability*, a term which refers to the totality of a person's mental capabilities. Psychologists know full well that what they mean by intelligence in the technical sense is only a part of the whole spectrum of human abilities. The notion that a person's intelligence, or some test measurement thereof, reflects the totality of all that he can possibly do with his "brains" has long caused much misunderstanding and needless dispute. As I have already indicated, the particular constellation of abilities we now call "intelligence," and which we can measure by means of "intelligence" tests, has been singled out from the total galaxy of mental abilities as being especially important in our society mainly because of the nature of our traditional system of formal education and the occupational structure with which it is coordinated. Thus, the predominant importance of intelligence is derived, not from any absolute criteria or God-given desiderata, but from societal demands. But neither does this mean, as some persons would like to believe, that intelligence exists only "by definition" or is merely an insubstantial figment of psychological theory and test construction. Intelligence fully meets the usual scientific criteria for being regarded as an aspect of objective reality, just as much as do atoms, genes, and electromagnetic fields. Intelligence has indeed been singled out as especially important by the educational and occupational demands prevailing in all industrial societies, but it is nevertheless a biological reality and not just a figment of social

convention. Where educators and society in general are most apt to go wrong is in failing fully to recognize and fully to utilize a broader spectrum of abilities than just that portion which psychologists have technically designated as "intelligence." But keep in mind that it is this technical meaning of "intelligence" to which the term specifically refers throughout the present article.

The Distribution of Intelligence

Intelligence tests yield numerical scores or IQs (intelligence quotients) which are assumed to be, and in fact nearly are, "normally" distributed in the population. That is, the distribution of IQs conforms to the normal or so-called Gaussian distribution, the familiar "bell-shaped curve." The IQ, which is now the most universal "unit" in the measurement of intelligence, was originally defined as the ratio of the individual's mental age (MA) to his chronological age (CA): $IQ = (MA/CA) \times 100$. (Beyond about 16 years of age, the formula ceases to make sense.) Mental age was simply defined as the typical or average score obtained on a test by children of a given age, and thus the average child by definition has an IQ of 100. Because of certain difficulties with the mental age concept, which we need not go into here, modern test constructors no longer attempt to measure mental age but instead convert raw scores (i.e., the number of test items gotten "right") directly into IQs for each chronological age group. The average IQ at each age is arbitrarily set at 100, and the IQ is defined as a normally distributed variable with a mean of 100 and a standard deviation of 15 points. (The standard deviation is an index of the amount of dispersion of scores; in the normal distribution 99.7 percent of the scores fall within $\pm$ 3 standard deviations [i.e., $\pm$ 45 IQ points] of the mean.)

There is really nothing mysterious about the fact that IQs are "normally" distributed, but it is not quite sufficient, either, to say that the normality of the distribution is just an artifact of test construction. There is a bit more to it than that.

Toss a hundred or so pennies into the air and record the number of heads that come "up" when they fall. Do this several thousand times and plot a frequency distribution of the number of heads that come up on each of the thousands of throws. You will have a distribution that very closely approximates the normal curve, and the more times you toss the hundred pennies the closer you will approximate the normal distribution.

Now, a psychological test made up of 100 or so items would behave in the same

manner as the pennies, and produce a perfectly normal distribution of scores, if (*a*) the items have an average difficulty level of 1/2 [i.e., exactly half of the number of persons taking the test would get the item "right"], and (*b*) the items are independent, that is, all the interitem correlations are zero. Needless to say, no psychological test that has ever been constructed meets these "ideal" criteria, and this is just as well, for if we succeeded in devising such a test it would "measure" absolutely nothing but chance variation. If the test is intended to measure some trait, such as general intelligence, it will be impossible for all the test items to be completely uncorrelated. They will necessarily have some degree of positive correlation among them. Then, if the items are correlated, and if we still want the test to spread people out over a considerable range of scores, we can achieve this only if the items vary in level of difficulty; they cannot all have a difficulty level of 1/2. (Imagine the extreme case in which all item intercorrelations were perfect and the difficulty level of all items was 1/2. Then the "distribution" of scores would have only two points: half the testees would obtain a score of zero and half would obtain a perfect score.) So we need to have test items which have an *average* difficulty level of 1/2 in the test overall, but which cover a considerable range of difficulty levels, say, from .1 to .9. Thus, test constructors make up their tests of items which have rather low average intercorrelations (usually between .1 and .2) and a considerable range of difficulty levels. These two sets of conditions working together, then, yield a distribution of test scores in the population which is very close to "normal." So far it appears as though we have simply made our tests in such a way as to *force* the scores to assume a normal distribution. And that is exactly true.

But the important question still remains to be answered: is intelligence itself—not just our measurements of it—really normally distributed? In this form the question is operationally meaningless, since, in order to find the form of the distribution of intelligence, we first have to measure it, and we have constructed our measuring instruments in such a way as to yield a normal distribution. The argument about the distribution of intelligence thus appears to be circular. Is there any way out? The only way I know of is to look for evidence that our intelligence scales or IQs behave like an "interval scale." On an interval scale, the interval between any two points is equal to the interval between any other two points the same numerical distance apart. Thus, intervals on the scale are equal and additive. If we *assume* that intelligence is "really" normally distributed in the population, and then measure it in such a way that we obtain a normal distri-

bution of scores, our measurements (IQs) can be regarded as constituting an interval scale. If, then, the scale in fact behaves like an interval scale, there is some justification for saying that intelligence itself (not just IQ) is normally distributed. What evidence is there of the IQ's behaving like an interval scale? The most compelling evidence, I believe, comes from studies of the inheritance of intelligence, in which we examine the pattern of intercorrelations among relatives of varying degrees of kinship.

But, first, to understand what is meant by "behaving" like an interval scale, let us look at two well-known interval scales, the Fahrenheit and centigrade thermometers. We can prove that these are true interval scales by showing that they "behave" like interval scales in the following manner: Mix a pint of ice water at 0° C with a pint of boiling water at 100° C. The resultant temperature of the mixture will be 50° C. Mix 3 pints of ice water with 1 pint of boiling water and the temperature of the mix will be 25° C. And we can continue in this way, mixing various proportions of water at different temperatures and predicting the resultant temperatures on the assumption of an interval scale. To the extent that the thermometer readings fit the predictions, they can be considered an interval scale.

Physical stature (height) is measured on an interval scale (more than that, it is also a ratio scale) in units which are independent of height, so the normal distribution of height in the population is clearly a fact of nature and not an artifact of the scale of measurement. A rather simple genetic model "explains" the distribution of height by hypothesizing that individual variations in height are the result of a large number of independent factors each having a small effect in determining stature. (Recall the penny-tossing analogy.) This model predicts quite precisely the amount of "regression to the population mean" of the children's average height from the parent's average height, a phenomenon first noted by Sir Francis Galton in 1885. The amount of "regression to the mean" from grandparent to grandchild is exactly double that from parent to child. These regression lines for various degrees of kinship are perfectly rectilinear throughout the entire range, except at the very lower end of the scale of height, where one finds midgets and dwarfs. The slope of the regression line changes in discrete jumps according to the remoteness of kinship of the groups being compared. All this could happen only if height were measured on an interval scale. The regression lines would not be rectilinear if the trait (height) were not measured in equal intervals.

Now, it is interesting that intelligence measurements show about the same degree of "filial regression," as Galton called it, that we find for height. The simple

77-943 O - 73 - 7

polygenic model for the inheritance of height fits the kinship correlations obtained for intelligence almost as precisely as it does for height. And the kinship regression lines are as rectilinear for intelligence as for height, throughout the IQ scale, except at the very lower end, where we find pathological types of mental deficiency analogous to midgets and dwarfs on the scale of physical stature. In brief, IQs behave just about as much like an interval scale as do measurements of height, which we know for sure is an interval scale. Therefore, it is not unreasonable to treat the IQ as an interval scale.

Although standardized tests such as the Stanford-Binet and the Wechsler Scales were each constructed by somewhat different approaches to achieving interval scales, they both agree in revealing certain systematic discrepancies from a perfectly normal distribution of IQs when the tests are administered to a very large and truly random sample of the population. These slight deviations of the distribution of IQs from perfect normality have shown up in many studies using a variety of tests. The most thorough studies and sophisticated discussions of their significance can be found in articles by Sir Cyril Burt (1957, 1963). The evidence, in short, indicates that intelligence is *not* distributed quite normally in the population. The distribution of IQs approximates normality quite closely in the IQ range from about 70 to 130. But outside this range there are slight, although very significant, departures from normality. From a scientific standpoint, these discrepancies are of considerable interest as genuine phenomena needing explanation.

Figure I shows an idealized distribution of IQs if they were distributed perfectly normally. Between IQ 70 and IQ 130, the percentage of cases falling between different IQ intervals, as indicated in Figure 1, are very close to the actual percentages estimated from large samples of the population and the departures are hardly enough to matter from any practical standpoint.

Examination of this normal curve can be instructive if one notes the consequences of shifting the total distribution curve up or down the IQ scale. The consequences of a given shift become more extreme out toward the "tails" of the distribution. For example, shifting the mean of the distribution from 100 down to 90 would put 50 percent instead of only 25 percent of the population below IQ 90; and it would put 9 percent instead of 2 percent below IQ 70. And in the upper tail of the distribution, of course, the consequences would be the reverse; instead of 25 percent above IQ 110, there would be only 9 percent, and so on. The point is that relatively small shifts in the mean of the IQ distribution can result in very large differences in the proportions of the population that fall into

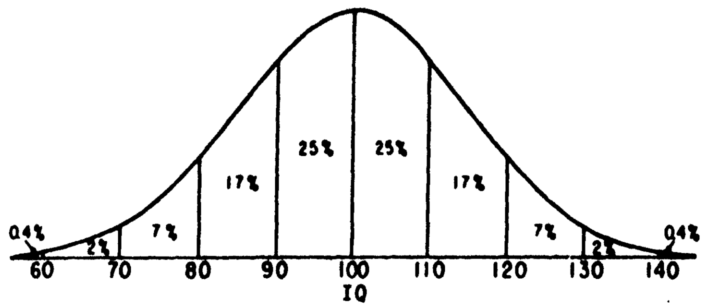


FIGURE 1.
The theoretical normal or Gaussian distribution of IQs, showing the expected percentages of the population in each IQ range. Except at the extremes (below 70 and above 130) these percentages are very close to actual population values. (The percentage figures total slightly more than 100% because of rounding.)

the very low or the very high ranges of intelligence. A 10 point downward shift in the mean, for example, would more than triple the percentage of mentally retarded (IQs below 70) in the population and would reduce the percentage of intellectually "gifted" (IQs above 130) to less than one-sixth of their present number. It is in these tails of the normal distribution that differences become most conspicuous between various groups in the population that show mean IQ differences, for whatever reason, of only a few IQ points. From a knowledge of relatively slight mean differences between various social class and ethnic groups, for example, one can estimate quite closely the relatively large differences in their proportions in special classes for the educationally retarded and for the "gifted" and in the percentages of different groups receiving scholastic honors at graduation. It is simply a property of the normal distribution that the effects of group differences in the mean are greatly magnified in the different proportions of each group that we find as we move further out toward the upper or lower extremes of the distribution.

I indicated previously that the distribution of intelligence is really not quite "normal," but shows certain systematic departures from "normality." These departures from the normal distribution are shown in Figure 2 in a slightly exaggerated form to make them clear. The shaded area is the normal distribution; the heavy line indicates the actual distribution of IQs in the population. We note that there are more very low IQs than would be expected in a truly normal distribution,

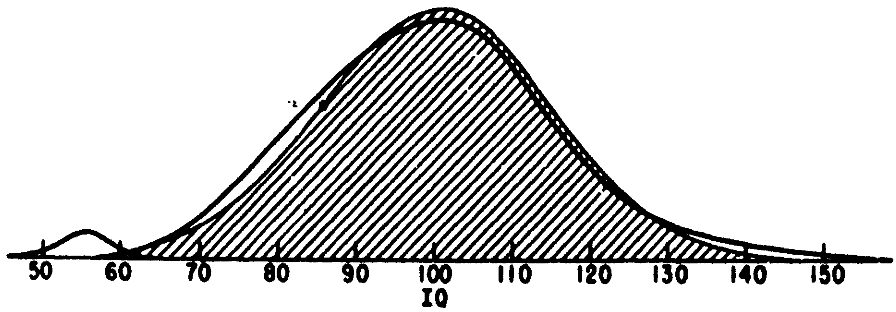


FIGURE 2.

Theoretical "normal" distribution of IQs (shaded curve) and the actual distribution in the population (heavy line), with the lower hump exaggerated for explanatory purposes. See text for explanation.

and also there is an excess of IQs at the upper end of the scale. Note, too, the slight excess in the IQ range between about 70 and 90.

The very lowest IQs, below 55 or 60, we now know, really represent a different distribution from that of the rest of the intelligence distribution (Roberts, 1952; Zigler, 1967). Whatever factors are responsible for individual differences in the IQ range above 60 are not sufficient to account for IQs below this level, and especially below IQ 50. Practically all IQs below this level represent severe mental deficiency due to pathological conditions, massive brain damage, or rare genetic and chromosomal abnormalities. Only about 1/2 to 3/4 of 1 percent of the total population falls into the IQ range below 50; this is fewer than 1/3 of all individuals classed as mentally retarded (IQs below 70). These severe grades of mental defect are not just the lower extreme of normal variation. Often they are due to a single recessive or mutant gene whose effects completely override all the other genetic factors involved in intelligence; thus they have been called "major gene" defects. In this respect, the distribution of intelligence is directly analogous to the distribution of stature. Short persons are no more abnormal than are average or tall persons; all are instances of normal variation. But extremely short persons at the very lower end of the distribution are really part of another, abnormal, distribution, generally consisting of midgets and dwarfs. They are clearly not a part of normal variation. One of the commonest types of dwarfism, for example, is known to be caused by a single recessive gene.

Persons with low IQs caused by major gene defects or chromosomal abnormal-

ities, like mongolism, are also usually abnormal in physical appearance. Persons with moderately low IQs that represent a part of normal variation, the so-called "familial mentally retarded," on the other hand, are physically indistinguishable from persons in the higher ranges of IQ. But probably the strongest evidence we have that IQs below 50 are a group apart from the mildly retarded, who represent the lower end of normal variation, comes from comparisons of the siblings of the severely retarded with siblings of the mildly retarded. In England, where this has been studied intensively, these two retardate groups are called imbecile (IQs below 50) and feebleminded (IQs 50 to 75). Figure 3 shows the IQ distributions of the *siblings* of imbecile and feebleminded children (Roberts, 1952). Note that the siblings of imbeciles have a much higher average level of intelligence than the siblings of the feebleminded. The latter group, furthermore, shows a distribution of IQs that would be predicted from a genetic model intended to account for the normal variation of IQ in the population. This model does not at all predict the IQ distribution for the imbecile sibships. To explain the results shown in Figure 3 one must postulate some additional factors (gene or chromosome defects, pathological conditions, etc.) that cause imbecile and idiot grades of mental deficiency.

Another interesting point of contrast between severe mental deficiency and mild retardation is the fact noted by Kushlick (1966, p. 130), in surveying numerous studies, that "The parents of severely subnormal children are evenly distributed among all the social strata of industrial society, while those of mildly subnormal subjects come predominantly from the lower social classes. There is

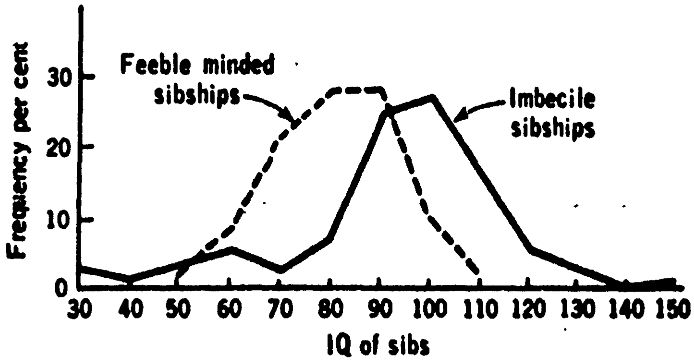


FIGURE 3.
Frequency distributions of the IQs of sibs of feebleminded and imbeciles of the IQ range 30-68. (Roberts, 1952.)

now evidence which suggests that mild subnormality in the absence of abnormal neurological signs (epilepsy, electroencephalographic abnormalities, biochemical abnormalities, chromosomal abnormalities or sensory defects) is virtually confined to the lower social classes. Indeed, there is evidence that almost no children of higher social class parents have IQ scores of less than 80, unless they have one of the pathological processes mentioned above."

In the remainder of this article we shall not be further concerned with these exceptionally low IQs below 50 or 60, which largely constitute a distribution of abnormal conditions superimposed on the factors that make for normal variation in intelligence. We shall be mainly concerned with the factors involved in the normal distribution.

Returning to Figure 2, the best explanation we have for the "bulge" between 70 and 90 is the combined effects of severe environmental disadvantages and of emotional disturbances that depress test scores. Burt (1963) has found that when, independent of the subjects' test performance there is evidence for the existence of factors that depress performance, and these exceptional subjects' scores are removed from the distribution, this "bulge" in the 70-90 range is diminished or erased. Also, on retest under more favorable conditions, the IQs of many of these exceptional subjects are redistributed at various higher points on the scale, thereby making the IQ distribution more normal.

The "excess" of IQs at the high end of the scale is certainly a substantial phenomenon, but it has not yet been adequately accounted for. In his multifactorial theory of the inheritance of intelligence, Burt (1958) has postulated major gene effects that make for exceptional intellectual abilities represented at the upper end of the scale, just as other major gene effects make for the subnormality found at the extreme lower end of the scale. One might also hypothesize that superior genotypes for intellectual development are pushed to still greater superiority in their phenotypic expression through interaction with the environment. Early recognition of superiority leads to its greater cultivation and encouragement by the individual's social environment. This influence is keenly evident in the developmental histories of persons who have achieved exceptional eminence (Goertzel & Goertzel, 1962). Still another possible explanation of the upper-end "excess" lies in the effects of assortative mating in the population, meaning the tendency for "like to marry like." If the degree of resemblance in intelligence between parents in the upper half of the IQ distribution were significantly greater than the degree of resemblance of parents in the below-average range, genetic theory would predict the relative elongation of the upper tail of the distribution.

This explanation, however, must remain speculative until we have more definite evidence of whether there is differential assortative mating in different regions of the IQ distribution.

The Concept of Variance. Before going on to discuss the factors that account for normal variation in intelligence among individuals in the population, a word of explanation is in order concerning the quantification of variation. The amount of dispersion of scores depicted by the distributions in Figures 1 and 2 is technically expressed as the *variance,* which is the square of the standard deviation of the scores in the distribution. (Since the standard deviation of IQs in the population is 15, the total variance is 225.) *Variance* is a basic concept in all discussions of individual differences and population genetics. If you take the difference between every score and the mean of the total distribution, square each of these differences, sum them up, and divide the sum by the total number of scores, you have a quantity called the *variance*. It is an index of the total amount of variation among scores. Since variance represents variation on an additive scale, the total variance of a distribution of scores can be partitioned into a number of components, each one due to some factor which contributes a certain specifiable proportion of the variance, and all these variance components add up to the total variance. The mathematical technique for doing this, called "the analysis of variance," was invented by Sir Ronald Fisher, the British geneticist and statistician. It is one of the great achievements in the development of statistical methodology.

The Inheritance of Intelligence

"In the actual race of life, which is not to get ahead, but to get ahead of somebody, the chief determining factor is heredity." So said Edward L. Thorndike in 1905. Since then, the preponderance of evidence has proved him right, certainly as concerns those aspects of life in which intelligence plays an important part.

But one would get a quite different impression from reading most of the recent popular textbooks of psychology and education. Genetic factors in individual differences have usually been belittled, obscured, or denigrated, probably for reasons of interest mainly on historical, political, and idealogical grounds which we need not go into here. Some of the following quotations, each from different widely used texts in our field, give some indication of the basis for my complaint. "We can attribute no particular portion of intelligence to heredity and no particular portion to the environment." "The relative influence of heredity and envi-

ronment upon intelligence has been the topic of considerable investigations over the last half century. Actually the problem is incapable of solution since studies do not touch upon the problem of heredity and environment but simply upon the susceptibility of the content of a particular test to environmental influences." "Among people considered normal, the range of genetic variations is not very great." "Although at the present time practically all responsible workers in the field recognize that conclusive proof of the heritability of mental ability (where no organic or metabolic pathology is involved) is still lacking, the assumption that subnormality has a genetic basis continues to crop up in scientific studies." "There is no evidence that nature is more important than nurture. These two forces always operate together to determine the course of intellectual development." The import of such statements apparently filters up to high levels of policy-making, for we find a Commissioner of the U.S. Office of Education stating in a published speech that children ". . . all have similar potential at birth. The differences occur shortly thereafter." These quotations typify much of the current attitude toward heredity and environment that has prevailed in education in recent years. The belief in the almost infinite plasticity of intellect, the ostrich-like denial of biological factors in individual differences, and the slighting of the role of genetics in the study of intelligence can only hinder investigation and understanding of the conditions, processes, and limits through which the social environment influences human behavior.

But fortunately we are beginning to see some definite signs that this mistreatment of the genetic basis of intelligence by social scientists may be on the wane, and that a biosocial view of intellectual development more in accord with the evidence is gaining greater recognition. As Yale psychologist Edward Zigler (1968) has so well stated:

> Not only do I insist that we take the biological integrity of the organism seriously, but it is also my considered opinion that our nation has more to fear from unbridled environmentalists than they do from those who point to such integrity as one factor in the determination of development. It is the environmentalists who have been writing review after review in which genetics are ignored and the concept of capacity is treated as a dirty word. It is the environmentalists who have placed on the defensive any thinker who, perhaps impressed by the revolution in biological thought stemming from discoveries involving RNA-DNA phenomena, has had the temerity to suggest that certain behaviors may be in part the product of read-out mechanisms residing within the programmed organism. It is the unbridled environmentalist who emphasizes the plasticity of the intellect, that tells us one can change both the general rate of development and the configuration of intellectual

processes which can be referred to as the intellect, if we could only subject human beings to the proper technologies. In the educational realm, this has spelled itself out in the use of panaceas, gadgets, and gimmicks of the most questionable sort. It is the environmentalist who suggests to parents how easy it is to raise the child's IQ and who has prematurely led many to believe that the retarded could be made normal, and the normal made geniuses. It is the environmentalist who has argued for pressure-cooker schools, at what psychological cost, we do not yet know.

Most geneticists and students of human evolution have fully recognized the role of culture in shaping "human nature," but also they do not minimize the biological basis of diversity in human behavioral characteristics. Geneticist Theodosius Dobzhansky (1968, p. 554) has expressed this viewpoint in the broadest terms: "The trend of cultural evolution has been not toward making everybody have identical occupations but toward a more and more differentiated occupational structure. What would be the most adaptive response to this trend? Certainly nothing that would encourage genetic uniformity.... To argue that only environmental circumstances and training determine a person's behavior makes a travesty of democratic notions of individual choice, responsibility, and freedom."

Evidence from Studies of Selective Breeding

The many studies of selective breeding in various species of mammals provide conclusive evidence that many behavioral characteristics, just as most physical characteristics, can be manipulated by genetic selection (see Fuller & Thompson, 1962; Scott and Fuller, 1965). Rats, for example, have been bred for maze learning ability in many different laboratories. It makes little difference whether one refers to this ability as rat "intelligence," "learning ability" or some other term—we know that it is possible to breed selectively for whatever the factors are that make for speed of maze learning. To be sure, individual variation in this complex ability may be due to any combination of a number of characteristics involving sensory acuity, drive level, emotional stability, strength of innate turning preferences, brain chemistry, brain size, structure of neural connections, speed of synaptic transmission, or whatever. The point is that the molar behavior of learning to get through a maze efficiently without making errors (i.e., going up blind alleys) can be markedly influenced in later generations by selective breeding of the parent generations of rats who are either fast or slow ("maze bright" or "maze dull," to use the prevailing terminology in this research) in learning to get through the maze. Figure 4 shows the results of one such genetic selection experiment. They are quite typical; within only six generations

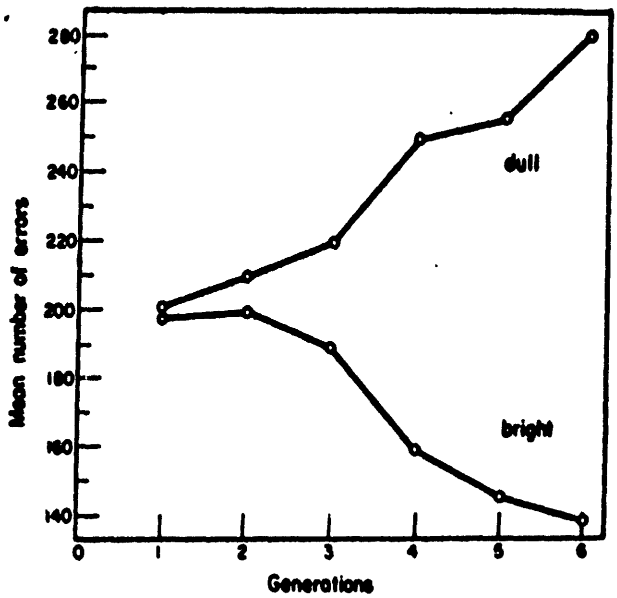


FIGURE 4.

The mean error scores in maze learning for successive generations of selectively bred "bright" and "dull" strains of McGill rats. (After Thompson, 1954.)

of selection the offspring of the "dull" strain make 100 percent more errors in learning the maze than do the offspring of the "bright" strain (Thompson, 1954). In most experiments of this type, of course, the behaviors that respond so dramatically to selection are relatively simple as compared with human intelligence, and the experimental selection pressure is severe, so the implications of such findings for the study of human variation should not be overdrawn. Yet geneticists seem to express little doubt that many behavioral traits in humans would respond similarly to genetic selection. Three eminent geneticists (James F. Crow, James V. Neel, and Curt Stern) of the National Academy of Sciences recently prepared a "position statement," which was generally hedged by extreme caution and understatement, that asserted: "Animal experiments have shown that almost any trait can be changed by selection. . . . A selection program to increase human intelligence (or whatever is measured by various kinds of 'intelligence' tests) would almost certainly be successful in some measure. The same is probably true for other behavioral traits. The *rate* of increase would be somewhat unpredictable, but there is little doubt that there would be progress" (National Academy of Sciences, 1967, p. 893).

Direct Evidence of Genetic Influences on Human Abilities

One of the most striking pieces of evidence for the genetic control of mental abilities is a chromosomal anomaly called Turner's syndrome. Normal persons have 46 chromosomes. Persons with Turner's syndrome have only 45. When their chromosomes are stained and viewed under the microscope, it is seen that the sex-chromatin is missing from one of the two chromosomes that determine the individual's sex. In normal persons this pair of chromosomes is conventionally designated XY for males and XX for females. The anomaly of Turner's syndrome is characterized as XO. These persons always have the morphologic appearance of females but are always sterile, and they show certain physical characteristics such as diminutive stature, averaging about five feet tall as adults. The interesting point about Turner's cases from our standpoint is that although their IQs on most verbal tests of intelligence show a perfectly normal distribution, their performance on tests involving spatial ability or perceptual organization is abnormally low (Money, 1964). Their peculiar deficiency in spatial-perceptual ability is sometimes so severe as to be popularly characterized as "space-form blindness." It is also interesting that Turner's cases seem to be more or less uniformly low on spatial ability regardless of their level of performance on other tests of mental ability. These rare persons also report unusual difficulty with arithmetic and mathematics in school despite otherwise normal or superior intelligence. So here is a genetic aberration, clearly identifiable under the microscope, which has quite specific consequences on cognitive processes. Such specific intellectual deficiencies are thus entirely possible without there being any specific environmental deprivations needed to account for them.

There are probably other more subtle cognitive effects associated with the sex chromosomes in normal persons. It has long been suspected that males have greater environmental vulnerability than females, and Nancy Bayley's important longitudinal research on children's mental development clearly shows both a higher degree and a greater variety of environmental and personality correlates of mental abilities in boys than in girls (Bayley, 1965b, 1966, 1968).

Polygenic Inheritance

Since intelligence is basically dependent on the structural and biochemical properties of the brain, it should not be surprising that differences in intellectual capacity are partly the result of genetic factors which conform to the same principles involved in the inheritance of physical characteristics. The general model that geneticists have devised to account for the facts of inheritance of

continuous or metrical physical traits, such as stature, cephalic index, and finger-print ridges, also applies to intelligence. *The mechanism of inheritance for such traits is called polygenic, since normal variation in the characteristic is the result of multiple genes whose effects are small, similar, and cumulative.* The genes can be thought of as the pennies in the coin-tossing analogy described previously. Some genes add a positive increment to the metric value of the characteristic ("heads") and some genes add nothing ("tails"). The random segregation of the parental genes in the process of gametogenesis (formation of the sex cells) and their chance combination in the zygote (fertilized egg) may be likened to the tossing of a large number of pennies, with each "head" adding a positive increment to the trait, thereby producing the normal bell-shaped distribution of trait values in a large number of tosses. The actual number of genes involved in intelligence is not known. In fact, the total number of genes in the human chromosomes is unknown. The simplest possible model would require between ten and twenty gene pairs (alleles) to account for the normal distribution of intelligence, but many more genes than this are most likely involved (Gottesman, 1963, pp. 290-291).

The Concept of Heritability

The study of the genetic basis of individual differences in intelligence in humans has evolved in the traditions and methods of that branch of genetics called quantitative genetics or population genetics, the foundations of which were laid down by British geneticists and statisticians such as Galton, Pearson, Fisher, Haldane, and Mather, and, in the United States, by J. L. Lush and Sewall Wright. Probably the most distinguished exponent of the application of these methods to the study of intelligence is Sir Cyril Burt, whose major writings on this subject are a "must" for students of individual differences (Burt, 1955, 1958, 1959, 1961, 1966; Burt & Howard, 1956, 1957).

One aim of this approach to the study of individual differences in intelligence is to account for the total variance in the population (excluding pathological cases at the bottom of the distribution) in terms of the proportions of the variance attributable to various genetic and environmental components. It will pay to be quite explicit about just what this actually means.

Individual differences in such measurements of intelligence as the IQ are represented as population variance in a phenotype V_P, and are distributed approximately as shown in Figure 1. Conceptually, this total variance of the phenotypes can be partitioned into a number of variance components, each of

which represents a source of variance. The components, of course, all add up to the total variance. Thus,

$$V_P = \underbrace{\frac{(V_G + V_{AM}) + V_D + V_i}{V_H}}_{\text{Heredity}} + \underbrace{\frac{V_E + 2\,Cov_{HE} + V_I}{V_E}}_{\text{Environment}} + \underbrace{\frac{V_e}{}}_{\text{Error}} \quad (2)$$

where: V_P = phenotypic variance in the population
V_G = genic (or additive) variance
V_{AM} = variance due to assortative mating. V_{AM} = O under random mating (panmixia).
V_D = dominance deviation variance
V_i = epistatis (interaction among genes at 2 or more loci)
V_E = environmental variance
Cov_{HE} = covariance of heredity and environment
V_I = true statistical interaction of genetic and environmental factors
V_e = error of measurement (unreliability).

Here are a few words of explanation about each of these variance components.

Phenotypic Variance. V_P is already clear; it is the total variance of the trait measurements in the population.

Genic Variance. V_G, the genic (or additive) variance, is attributable to gene effects which are additive; that is, each gene adds an equal increment to the metric value of the trait. Sir Ronald Fisher referred to this component as "the essential genotypes," since it is the part of the genetic inheritance which "breeds true"—it accounts for the resemblance between parents and offspring. If trait variance involved nothing but additive genic effects, the average value of all the offspring that could theoretically be born to a pair of parents would be exactly equal to the average value of the parents (called the midparent value). It is thus the genic aspect which is most important to agriculturalists and breeders of livestock, since it is the genic component of the phenotypic variance that responds to selection according to the simple rule of "like begets like." The larger the proportion of genic variance involved in a given characteristic, the fewer is the number of generations of selective breeding required to effect a change of some specified magnitude in the characteristic.

Assortative Mating. V_{AM}, the variance due to assortative mating, is conventionally not separated from V_G, since assortative mating actually affects the proportion

of V_G directly. I have separated these components here for explanatory reasons, and it is, in fact, possible to obtain independent estimates of the two components. If mating were completely random in the population with respect to a given characteristic—that is, if the correlation between parents were zero (a state of affairs known as *panmixia*)—the V_{AM} component would also be equal to zero and the population variance on the trait in question would therefore be reduced.

Assortative mating has the effect of increasing the resemblance among siblings and also of increasing the differences between families in the population. (In the terminology of analysis of variance, assortative mating decreases the *Within* Families variance and increases the *Between* Families variance.)

For some human characteristics the degree of assortative mating is effectively zero. This is true of fingerprint ridges, for example. Men and women are obviously not attracted to one another on the basis of their fingerprints. Height, however, has an assortative mating coefficient (i.e., the correlation between mates) of about .30. The IQ, interestingly enough, shows a higher degree of assortative mating in our society than any other measurable human characteristic. I have surveyed the literature on this point, based on studies in Europe and North America, and find that the correlation between spouses' intelligence test scores averages close to +.60. Thus, spouses are more alike in intelligence than brothers and sisters, who are correlated about .50.

As Eckland (1967) has pointed out, this high correlation between marriage partners does not come about solely because men and women are such excellent judges of one another's intelligence, but because mate selection is greatly aided by the highly visible selective processes of the educational system and the occupational hierarchy. Here is a striking instance of how educational and social factors can have far-reaching genetic consequences in the population. One would predict, for example, that in preliterate or preindustrial societies assortative mating with respect to intelligence would be markedly less than it is in modern industrial societies. The educational screening mechanisms and socioeconomic stratification by which intelligence becomes more readily visible would not exist, and other traits of more visible importance to the society would take precedence over intelligence as a basis for assortative mating. Even in our own society, there may well be differential degrees of assortative mating in different segments of the population, probably related to their opportunities for educational and occupational selection. When any large and socially insulated group is not subject to the social and educational circumstances that lead to a high degree of assortative mating for intelligence, there should be important genetic

consequences. One possible consequence is some reduction of the group's ability, not as individuals but as a group, to compete intellectually. Thus probably one of the most cogent arguments for society's promoting full equality of educational, occupational, and economic opportunity lies in the possible genetic consequences of these social institutions.

The reason is simply that assortative mating increases the genetic variance in the population. By itself this will not affect the mean of the trait in the population, but it will have a great effect on the proportion of the population falling in the upper and lower tails of the distribution. Under present conditions, with an assortative mating coefficient of about .60, the standard deviation of IQs is 15 points. If assortative mating for intelligence were reduced to zero, the standard deviation of IQs would fall to 12.9. The consequences of this reduction in the standard deviation would be most evident at the extremes of the intelligence distribution. For example, assuming a normal distribution of IQs and the present standard deviation of 15, the frequency (per million) of persons above IQ 130 is 22,750. Without assortative mating the frequency of IQs over 130 would fall to 9,900, or only 43.5 percent of the present frequency. For IQs above 145, the frequency (per million) is 1,350 and with no assortative mating would fall to 241, or 17.9 percent of the present frequency. And there are now approximately 20 times as many persons above an IQ of 160 as we would find if there were no assortative mating for intelligence.[8] Thus differences in assortative mating can have a profound effect on a people's intellectual resources, especially at the levels of intelligence required for complex problem solving, invention, and scientific and technological innovation.

But what is the effect of assortative mating on the lower tail of the distribution? On theoretical grounds we should also expect it to increase the proportion of low IQs in the population. It probably does this to some extent, but not as much as it increases the frequency of higher IQs, because there is a longer-term consequence of assortative mating which must also be considered. A number of studies have shown that in populations practicing a high degree of assortative mating, persons below IQ 75 are much less successful in finding marriage partners and, as a group, have relatively fewer offspring than do persons of higher intelligence (Bajema, 1963, 1966; Higgins, Reed, & Reed, 1962). Since assortative mating increases variance, it in effect pushes more people into the below IQ

[8] I am grateful to University of California geneticist Dr. Jack Lester King for making these calculations, which are based on the assumption that the heritability of IQ is .80, a value which is the average of all the major studies of the heritability of intelligence.

75 group, where they fail to reproduce, thereby resulting in a net selection for genes favoring high intelligence. Thus, in the long run, assortative mating may have a eugenic effect in improving the general level of intelligence in the population.

Dominance Deviation. V_D, the dominance deviation variance, is apparent when we observe a systematic discrepancy between the average value of the parents and the average value of their offspring on a given characteristic. Genes at some of the loci in the chromosome are recessive (r) and their effects are not manifested in the phenotype unless they are paired with another recessive at the same locus. If paired with a dominant gene (D), their effect is overridden or "dominated" by the dominant gene. Thus, in terms of increments which genes add to the metric value of the phenotype, if $r = 0$ and $D = 1$, then $r + r = 0$, and $D + D = 2$, but $D + r$ will equal 2, since D dominates r. Because of the presence of some proportion of recessive genes in the genotypes for a particular trait, not all of the parents' phenotypic characteristics will show up in their offspring, and, of course, vice versa: not all of the offspring's characteristics will be seen in the parents. This makes for a less than perfect correlation between midparent and midchild values on the trait in question. V_D, the dominance variance, represents the component of variance in the population which is due to this average discrepancy between parents and offspring. The magnitude of V_D depends upon the proportions of dominant and recessive genes constituting the genotypes for the characteristic in the population.

Epistasis. V_I is the variance component attributable to epistasis, which means the interaction of the effects among genes at two or more loci. When genes "interact," their effects are not strictly additive; that is to say, their combined effect may be more or less than the sum of their separate effects. Like dominance, epistasis also accounts for some of the lack of resemblance between parents and their offspring. And it increases the population variance by a component designated as V_I.

Environmental Variance. "Environmental" really means all sources of variance not attributable to genetic effects or errors of measurement (i.e., test unreliability). In discussions of intelligence, the environment is often thought of only in terms of the social and cultural influences on the individual. While these are important, they are not the whole of "environment," which includes other more strictly biological influences, such as the prenatal environment and nutritional

factors early in life. In most studies of the heritability of intelligence "environment" refers to all variance that is not accounted for by genetic factors $[(V_G + V_{AM}) + V_D + V_I]$ and measurement error (V_e).

Covariance of Heredity and Environment. This term can also be expressed as $2r_{HE}\sqrt{V_H \times V_E}$, where r_{HE} is the correlation between heredity and environment, V_H is the variance due to all genetic factors, and V_E is variance due to all environmental factors. In other words, if there is a positive correlation between genetic and environmental factors, the population variance is increased by a theoretically specifiable amount indicated by the covariance term in Equation 2.

Such covariance undoubtedly exists for intelligence in our society. Children with better than average genetic endowment for intelligence have a greater than chance likelihood of having parents of better than average intelligence who are capable of providing environmental advantages that foster intellectual development. Even among children within the same family, parents and teachers will often give special attention and opportunities to the child who displays exceptional abilities. A genotype for superior ability may cause the social environment to foster the ability, as when parents perceive unusual responsiveness to music in one of their children and therefore provide more opportunities for listening, music lessons, encouragement to practice, and so on. A bright child may also create a more intellectually stimulating environment for himself in terms of the kinds of activities that engage his interest and energy. And the social rewards that come to the individual who excels in some activity reinforce its further development. Thus the covariance term for any given trait will be affected to a significant degree by the kinds of behavioral propensities the culture rewards or punishes, encourages or discourages. For traits viewed as desirable in our culture, such as intelligence, hereditary and environmental factors will be positively correlated. But for some other traits which are generally viewed as socially undesirable, hereditary and environmental influences may be negatively correlated. This means that the social environment tends to discourage certain behavioral propensities when they are out of line with the values of the culture. Then, instead of heredity and environment acting in the same direction, they work in opposite directions, with a consequent reduction in the population variance in the trait. Overt aggressive tendencies may be a good example of behavior involving a negative correlation between genotypic propensities and environmental counter-pressures. An example of negative heredity-environment

77-942 O - 72 - 8

correlation in the scholastic realm would be found in the case where a child with a poor genetic endowment for learning some skill which is demanded by societal norms, such as being able to read, causes the child's parents to lavish special tutorial attention on their child in an effort to bring his performance up to par.

In making overall estimates of the proportions of variance attributable to hereditary and environmental factors, there is some question as to whether the covariance component should be included on the side of heredity or environment. But there can be no "correct" answer to this question. To the degree that the individual's genetic propensities cause him to fashion his own environment, given the opportunity, the covariance (or some part of it) can be justifiably regarded as part of the total heritability of the trait. But if one wishes to estimate what the heritability of the trait would be under artificial conditions in which there is absolutely no freedom for variation in individuals' utilization of their environment, then the covariance term should be included on the side of environment. Since most estimates of the heritability of intelligence are intended to reflect the existing state of affairs, they usually include the covariance in the proportion of variance due to heredity.

Interaction of Heredity and Environment. The *interaction* of genetic and environmental factors (V_I) must be clearly distinguished from the *covariance* of heredity and environment. There is considerable confusion concerning the meaning of interaction in much of the literature on heredity and intelligence. It is claimed, for example, that nothing can be said about the relative importance of heredity and environment because intelligence is the result of the "interaction" of these influences and therefore their independent effects cannot be estimated. This is simply false. The proportion of the population variance due to genetic $\times$ environment interaction is conceptually and empirically separable from other variance components, and its independent contribution to the total variance can be known. Those who call themselves "interactionists," with the conviction that they have thereby either solved or risen above the whole issue of the relative contributions of heredity and environment to individual differences in intelligence, are apparently unaware that the preponderance of evidence indicates that the interaction variance, V_I is the smallest component of the total phenotypic variance of intelligence.

What *interaction* really means is that different genotypes respond in different ways to the same environmental factors. For example, genetically different individuals having the same initial weight and the same activity level may gain

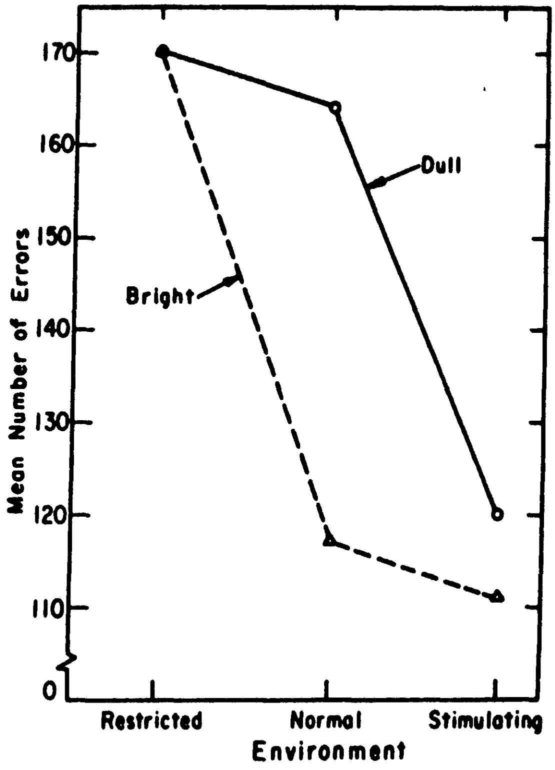


FIGURE 5.

Illustration of a true genotype × environment interaction for error scores in maze learning by "bright" and "dull" strains of rats raised in "restricted," "normal," and "stimulating" environments. (After Cooper & Zubek, 1958.)

weight at quite different rates all under exactly the same increase in caloric intake. Their genetically different constitutions cause them to metabolize exactly the same intake quite differently. An example of genotype × environmental interaction in the behavioral realm is illustrated in Figure 5. Strains of rats selectively bred for "brightness" or "dullness" in maze learning show marked differences in maze performance according to the degree of sensory stimulation in the conditions under which they are reared. For the "bright" strain, the difference between being reared in a "restricted" or in a "normal" environment makes a great difference in maze performance. But for the "dull" strain the

big difference is between a "normal" and a "stimulating" environment. While the strains differ greatly when reared under "normal" conditions (presumably the conditions under which they were selectively bred for "dullness" and "brightness"), they do not differ in the least when reared in a "restricted" environment and only slightly in a "stimulating" environment. This is the meaning of the genetic $\times$ environment interaction. Criticisms of the analysis of variance model for the components of phenotypic variance (e.g., Equation 2), put forth first by Loevinger (1943) and then by Hunt (1961, p. 329), are based on the misconception that the model implies that all effects of heredity and environment are strictly additive and there is no "non-additive" or interaction term. The presence of V_I in Equation 2 explicitly shows that the heredity $\times$ environment interaction is included in the analysis of variance model, and the contribution of V_I to the total variance may be estimated independently of the purely additive effects of heredity and environment. The magnitude of V_I for any given characteristic in any specified population is a matter for empirical study, not philosophic debate. If V_I turns out to constitute a relatively small proportion of the total variance, as the evidence shows is the case for human intelligence, this is not a fault of the analysis of variance model. It is simply a fact. If the interaction variance actually exists in any significant amount, the model will reveal it.

Several studies, reviewed by Wiseman (1964, p. 55; 1966, p. 66), provide most of the information we have concerning what may be presumed to be an heredity $\times$ environment interaction with respect to human intelligence. The general finding is that children who are more than one standard deviation (SD) above the mean IQ show greater correlations with environmental factors than do children who are more than one SD below the mean. In other words, if the heritability of IQ were determined in these two groups separately, it would be higher in the low IQ groups. Also, when siblings within the same family are grouped into above and below IQ 100, the scholastic achievement of the above 100 group shows a markedly higher correlation with environmental factors than in the below 100 group. This indicates a true interaction between intelligence and environment in determining educational attainments.

Error Variance. The variance due to errors of measurement (V_e) is, of course, unwanted but unavoidable, since all measurements fall short of perfect reliability. The proportion of test score variance due to error is equal to $1-r_{tt}$ (where r_{tt} is the reliability of the test, that is, its correlation with itself). For most intelligence tests, error accounts for between 5 and 10 percent of the variance.

Definition of Heritability

Heritability is a technical term in genetics meaning specifically the proportion of phenotypic variance due to variance in genotypes. When psychologists speak of heritability they almost invariably define it as:

$$H = \frac{(V_G + V_{AM}) + V_D + V_I}{V_P - V_e} \qquad (3)$$

Although this formula is technically the definition of *H*, heritability estimates in psychological studies may also include the covariance term of Equation 2 in the numerator of Equation 3.

Common Misconceptions About Heritability

Certain misconceptions about heritability have become so widespread and strongly ingrained that it is always necessary to counteract them before presenting the empirical findings on the subject, lest these findings only add to the confusion or provoke the dogmatic acceptance or rejection of notions that are not at all implied by the meaning of heritability.

Heredity versus Environment. Genetic and environmental factors are not properly viewed as being in opposition to each other. Nor are they an "all or none" affair. Any observable characteristic, physical or behavioral, is a phenotype, the very existence of which depends upon both genetic and environmental conditions. The legitimate question is not whether the characteristic is due to heredity *or* environment, but what proportion of the population variation in the characteristic is attributable to genotypic variation (which is *H*, the heritability) and what proportion is attributable to non-genetic or environmental variation in the population (which is 1-*H*). For metric characteristics like stature and intelligence, *H* can have values between 0 and 1.

Individual versus Population. Heritability is a population statistic, describing the relative magnitude of the genetic component (or set of genetic components) in the population variance of the characteristic in question. It has no sensible meaning with reference to a measurement or characteristic in an individual. A single measurement, by definition, has no variance. There is no way of partitioning a given individual's IQ into hereditary and environmental components, as if the person inherited, say, 80 points of IQ and acquired 20 additional points from his environment. This is, of course, nonsense. *The square root of the heritability* ($\sqrt{H}$)*, however, tells us the correlation between genotypes and*

phenotypes in the population, and this permits a probabalistic inference concerning the average amount of difference between individuals' obtained IQs and the "genotypic value" of their intelligence. (The average correlation between phenotypes and genotypes for IQ is about .90 in European and North American Caucasian populations, as determined from summary data presented later in this paper [Table 2]. The square of this value is known as the heritability—the proportion of phenotypic variance due to genetic variation.) The principle is the same as estimating the "true" scores from obtained scores in test theory. Statements about individuals can be made only on a probabilistic basis and not with absolute certainty. Only if heritability were unity (i.e., $H = 1$) would there be a perfect correlation between obtained scores and genotypic values, in which case we could say with assurance that an individual's measured IQ perfectly represented his genotype for intelligence. This still would not mean that the phenotype could have developed without an environment, for without either heredity or environment there simply is no organism and no phenotype. Thus the statement we so often hear in discussions of individual differences—that the individual's intelligence is the product of the interaction of his heredity and his environment—is rather fatuous. It really states nothing more than the fact that the individual exists.

Constancy. From what has already been said about heritability, it must be clear that it is not a constant like π and the speed of light. H is an empirically determined population statistic, and like any statistic, its value is affected by the characteristics of the population. H will be higher in a population in which environmental variation relevant to the trait in question is small, than in a population in which there is great environmental variation. Similarly, when a population is relatively homogeneous in genetic factors but not in the environmental factors relevant to the development of the characteristic, the heritability of the characteristic in question will be lower. In short, the value of H is jointly a function of genetic and environmental variability in the population. Also, like any other statistic, it is an estimate based on a sample of the population and is therefore subject to sampling error—the smaller the sample, the greater the margin of probable error. Values of H reported in the literature do not represent what the heritability might be under any environmental conditions or in all populations or even in the same population at different times. Estimates of H are specific to the population sampled, the point in time, how the measurements were made, and the particular test used to obtain the measurements.

Measurements versus Reality. It is frequently argued that since we cannot really measure intelligence we cannot possibly determine its heritability. Whether we can or cannot measure intelligence, which is a separate issue I have already discussed, let it be emphasized that it makes no difference to the question of heritability. We do not estimate the heritability of some trait that lies hidden behind our measurements. We estimate the heritability of the phenotypes and these are the measurements themselves. Regardless of what it is that our tests measure, the heritability tells us how much of the variance in these measurements is due to genetic factors. If the test scores get at nothing genetic, the result will simply be that estimates of their heritability will not differ significantly from zero. The fact that heritability estimates based on IQs differ very significantly from zero is proof that genetic factors play a part in individual differences in IQ. To the extent that a test is not "culture-free" or "culture-fair," it will result in a lower heritability measurement. It makes no more sense to say that intelligence tests do not really measure intelligence but only *developed* intelligence than to say that scales do not really measure a person's weight but only the weight he has acquired by eating. An "environment-free" test of intelligence makes as much sense as a "nutrition-free" scale for weight.

Know All versus Know Nothing. This expression describes another confused notion: the idea that unless we can know absolutely *everything* about the genetics of intelligence we can know nothing! Proponents of this view demand that we be able to spell out in detail every single link in the chain of causality from genes (or DNA molecules) to test scores if we are to say anything about the heritability of intelligence. Determining the heritability of a characteristic does not at all depend upon a knowledge of its physical, biochemical, or physiological basis or of the precise mechanisms through which the characteristic is modified by the environment. Knowledge of these factors is, of course, important in its own right, but we need not have such knowledge to establish the genetic basis of the characteristic. Selective breeding was practiced fruitfully for centuries before anything at all was known of chromosomes and genes, and the science of quantitative genetics upon which the estimation of heritability depends has proven its value independently of advances in biochemical and physiological genetics.

Acquired versus Inherited. How can a socially defined attribute such as intelligence be said to be inherited? Or something that is so obviously acquired from the social environment as vocabulary? Strictly speaking, of course, only genes are inherited. But the brain mechanisms which are involved in learning are gene-

tically conditioned just as are other structures and functions of the organism. What the organism is capable of learning from the environment and its rate of learning thus have a biological basis. Individuals differ markedly in the amount, rate, and kinds of learning they evince even given equal opportunities. Consider the differences that show up when a Mozart and the average run of children are given music lessons! If a test of vocabulary shows high heritability, it only means that persons in the population have had fairly equal opportunity for learning all the words in the test, and the differences in their scores are due mostly to differences in capacity for learning. If members of the population had had very unequal exposures to the words in the vocabulary test, the heritability of the scores would be very low.

Immutability. High heritability by itself does not necessarily imply that the characteristic is immutable. Under greatly changed environmental conditions, the heritability may have some other value, or it may remain the same while the mean of the population changes. At one time tuberculosis had a very high heritability, the reason being that the tuberculosis bacilli were extremely widespread throughout the population, so that the main factor determining whether an individual contracted tuberculosis was not the probability of exposure but the individual's inherited physical constitution. Now that tuberculosis bacilli are relatively rare, difference in exposure rather than in physical predisposition is a more important determinant of who contracts tuberculosis. In the absence of exposure, individual differences in predisposition are of no consequence.

Heritability also tells us something about the locus of control of a characteristic. The control of highly heritable characteristics is usually in the organism's internal biochemical mechanisms. Traits of low heritability are usually controlled by external environmental factors. No amount of psychotherapy, tutoring, or other psychological intervention will elicit normal performance from a child who is mentally retarded because of phenylketonuria (PKU), a recessive genetic defect of metabolism which results in brain damage. Yet a child who has inherited the genes for PKU can grow up normally if his diet is controlled to eliminate certain proteins which contain phenylalanine. Knowledge of the genetic and metabolic basis of this condition in recent years has saved many children from mental retardation.

Parent-Child Resemblance. The old maxim that "like begets like" is held up as an instance of the workings of heredity. The lack of parent-child resemblance, n the other hand, is often mistakenly interpreted as evidence that a character-

istic is not highly heritable. But the principles of genetics also explain the fact that often "like begets unlike." A high degree of parent-offspring resemblance, in fact, is to be expected only in highly inbred (or homozygous) strains, as in certain highly selected breeds of dogs and laboratory strains of mice. The random segregation of the parental genes in the formation of the sex cells means that the child receives a random selection of only half of each parent's genes. This fact that parent and child have only 50 percent of their genes in common, along with the effects of dominance and epistasis, insures considerable genetic dissimilarity between parent and child as well as among siblings, who also have only 50 percent of their genes in common. The fact that one parent and a child have only 50 percent of their genes in common is reflected in the average parent-offspring correlation (r_{po}) of between .50 and .60 (depending on the degree of assortative mating for a given characteristic) which obtains for height, head circumference, fingerprint ridges, intelligence, and other highly heritable characteristics. (The correlation is also between .50 and .60 for siblings on these characteristics; sibling resemblance is generally much *higher* than this for traits of *low* heritability.) The genetic correlation between the average of both parents (called the "midparent") and a single offspring ($r_{\bar{p}o}$) is the square root of the correlation for a single parent (i.e., $r_{\bar{p}o} = \sqrt{r_{po}}$). The correlation between the average of *both* parents and the average of *all* the offspring ("midchild") that they could theoretically produce ($r_{\bar{p}\bar{o}}$) is the same value as H_N, i.e., heritability in the narrow sense.[4] It is noteworthy that empirical determinations of the midparent-midchild correlation ($r_{\bar{p}\bar{o}}$) in fact closely approximate the values of H as estimated by various methods, such as comparisons of twins, siblings, and unrelated children reared together.

Empirical Findings on the Heritability of Intelligence

It is always preferable, of course, to have estimates of the proportions of variance contributed by each of the components in Equation 2 than to have merely an overall estimate of H. But to obtain reliable estimates of the separate components requires large samples of persons of different kinships, such as identical twins reared together and reared apart, fraternal twins, siblings, half-siblings, parents-children,

[4] Heritability in the narrow sense is an estimate of the proportion of genic variance without consideration of dominance and epistasis. This contrasts with equation (3), the definition of H, which includes estimates for these two factors. Signified as H_N, heritability in the narrow sense is conceptually defined as:

$$H_N = \frac{(V_G + V_{AM})}{V_p - V_e}$$

cousins, and so on. The methods of quantitative genetics by which these variance components, as well as the heritability, can be calculated from such kinship data are technical matters beyond the scope of this article, and the reader must be referred elsewhere for expositions of the methodology of quantitative genetics (Cattell, 1960; Falconer, 1960; Huntley, 1966; Kempthorne, 1957; Loehlin, in press).

The most satisfactory attempt to estimate the separate variance components is the work of Sir Cyril Burt (1955, 1958), based on large samples of many kinships drawn mostly from the school population of London. The IQ test used by Burt was an English adaptation of the Stanford-Binet. Burt's results may be regarded as representative of variance components of intelligence in populations that are similar to the population of London in their degree of genetic heterogenity and in their range of environmental variation. Table 1 shows the percentage of variance due to the various components, grouped under "genetic" and "environmental," in Burt's analysis.

TABLE 1

Analysis of Variance of Intelligence Test Scores (Burt, 1958)

Source of Variance	Percent*	
Genetic:		
Genic (additive)	40.5	(47.9)
Assortative Mating	19.9	(17.9)
Dominance & Epistasis	16.7	(21.7)
Environmental:		
Covariance of Heredity & Environment	10.6	(1.4)
Random Environmental Effects, including $H \times E$ interaction (V_I)	5.9	(5.8)
Unreliability (test error)	6.4	(5.3)
Total	100.0	(100.0)

* Figures in parentheses are percentages for adjusted assessments. See text for explanation.

When Burt submitted the test scores to the children's teachers for criticism on the basis of their impressions of the child's "brightness," a number of children were identified for whom the IQ was not a fair estimate of the child's ability in the teacher's judgment. These children were retested, often on a number of tests on several occasions, and the result was an "adjusted" assessment of the child's

IQ. The results of the analysis of variance after these adjusted assessments were made are shown in parentheses in Table 1. Note that the component most affected by the adjustments is the covariance of heredity and environment, which is what we should expect if the test is not perfectly "culture-fair." It means that the adjusted scores reduced systematic environmental sources of variance and thereby came closer to representing the children's innate ability, or, stated more technically, the adjusted scores increased the correlation between genotype and phenotype from .88 for unadjusted scores to .93 for adjusted scores. (Corrected for test unreliability these correlations become .90 and .96, respectively. And the heritabilities (H_B) for the two sets of scores are therefore $(.90)^2 = .81$ and $(.96)^2 = .93$, respectively.)

Kinship Correlations. The basic data from which variance components and heritability coefficients are estimated are correlations among individuals of different degrees of kinship. Nearly all such kinship correlations reported in the literature are summarized in Table 2. The median values of the correlations obtained in the various studies are given here. These represent the most reliable values we have for the correlations among relatives. Most of the values are taken from the survey by Erlenmeyer-Kimling and Jarvik (1963), and I have supplemented these with certain kinship correlations not included in their survey and reported in the literature since their review (e.g., Burt, 1966, p. 150). The Erlenmeyer-Kimling and Jarvik (1963) review was based on 52 independent studies of the correlations of relatives for tested intellectual abilities, involving over 30,000 correlational pairings from 8 countries in 4 continents, obtained over a period of more than two generations. The correlations were based on a wide variety of mental tests, administered under a variety of conditions by numerous investigators with contrasting views regarding the importance of heredity. The authors conclude: "Against this pronounced heterogeneity, which should have clouded the picture, and is reflected by the wide range of correlations, a clearly definite consistency emerges from the data. The composite data are compatible with the polygenic hypothesis which is generally favored in accounting for inherited differences in mental ability" (Erlenmeyer-Kimling & Jarvik, 1963, p. 1479).

The compatibility with the polygenic hypothesis to which the authors (as outlined earlier on p. 53) refer can be appreciated in Table 2 by comparing the median values of the obtained correlations with the sets of theoretical values shown in the last two columns. The first set (Theoretical Value[1]) is based on calculations by Burt (1966), using the methods devised by Fisher for estimating

TABLE 2

Correlations for Intellectual Ability: Obtained and Theoretical Values

Correlations Between	*Number of Studies*	*Obtained Median r*[*]	*Theoretical Value*[1]	*Theoretical Value*[2]
Unrelated Persons				
Children reared apart	4	−.01	.00	.00
Foster parent and child	3	+.20	.00	.00
Children reared together	5	+.24	.00	.00
Collaterals				
Second Cousins	1	+.16	+ .14	+ .063
First Cousins	3	+.26	+ .18	+ .125
Uncle (or aunt) and nephew (or niece)	1	+.34	+ .31	+ .25
Siblings, reared apart	33	+.47	+ .52	+ .50
Siblings, reared together	36	+.55	+ .52	+ .50
Dizygotic twins, different sex	9	+.49	+ .50	+ .50
Dizygotic twins, same sex	11	+.56	+ .54	+ .50
Monozygotic twins, reared apart	4	+.75	+1.00	+1.00
Monozygotic twins, reared together	14	+.87	+1.00	+1.00
Direct Line				
Grandparent and grandchild	3	+.27	+ .31	+ .25
Parent (as adult) and child	13	+.50	+ .49	+ .50
Parent (as child) and child	1	+.56	+ .49	+ .50

[*] Correlations not corrected for attenuation (unreliability).
[1] Assuming assortative mating and partial dominance.
[2] Assuming random mating and only additive genes, i.e., the simplest possible polygenic model.

kinship correlations for physical characteristics involving assortative mating and some degree of dominance. The second set (Theoretical Value[2]) of theoretical values is based on the simplest possible polygenic model, assuming random mating and nothing but additive gene effects. So these are the values one would expect if genetic factors alone were operating and the trait variance reflected no environmental influences whatsoever.

First of all, one can note certain systematic departures of the obtained correlations from the theoretical values. These departures are presumably due to nongenetic or environmental influences. The orderly nature of these environmental effects, as reflected in the Erlenmeyer-Kimling and Jarvik median correlations, can be highlighted by graphical presentation, as shown in Figure 6. Note that the condition of being reared together or reared apart has the same effect on the difference in magnitudes of the correlations for the various kinships. (The

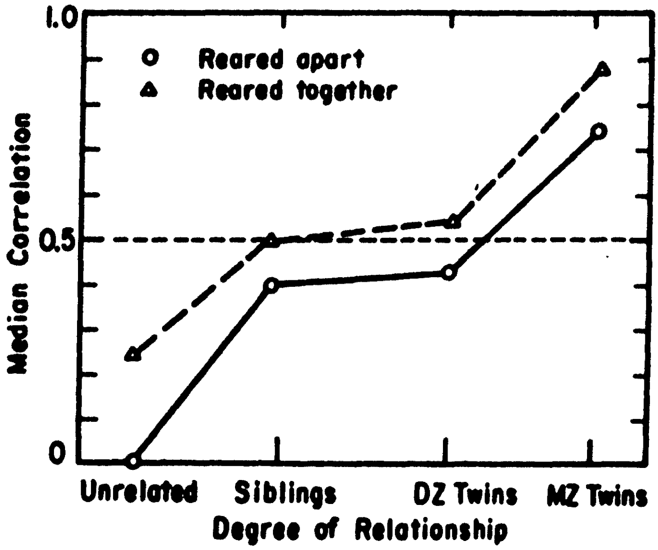


FIGURE 6.

Median values of all correlations reported in the literature up to 1963 for the indicated kinships. (After Erlenmeyer-Kimling & Jarvik, 1963.) Note consistency of difference in correlations for relatives reared together and reared apart.

slightly greater difference for unrelated children is probably due to the fact of selective placement by adoption agencies, that is, the attempt to match the child's intelligence with that of the adopting parents.)

Heritability Estimates. By making certain comparisons among the correlations shown in Table 2 and Figure 6, one can get some insight into how heritability is estimated. For example, we see that the correlation between identical or monozygotic (MZ) twins reared apart is .75. Since MZ twins develop from a single fertilized ovum and thus have exactly the same genes, any difference between the twins must be due to nongenetic factors. And if they are reared apart in uncorrelated environments, the difference between a perfect correlation (1.00) and the obtained correlation (.75) gives an estimate of the proportion of the variance in IQs attributable to environmental differences: $1.00 - 0.75 = 0.25$. Thus 75 percent of the variance can be said to be due to genetic variation (this is the heritability) and 25 percent to environmental variation. Now let us go to the other extreme and look at unrelated children reared together. They have no genetic inheritance in common, but they are reared in a common environment. Therefore the cor-

relation between such children will reflect the environment. As seen in Table 2, this correlation is 0.24. Thus, the proportion of IQ variance due to environment is .24; and the remainder, 1.00 — .24 = .76 is due to heredity. There is quite good agreement between the two estimates of heritability.

Another interesting comparison is between MZ twins reared together ($r = .87$) and reared apart ($r = .75$). If 1.00 — .75 = .25 (from MZ twins reared apart) estimates the total environmental variance, then 1.00 — .87 = .13 (from MZ twins reared together) is an estimate of the environmental variance *within families* in which children are reared together. Thus the difference between .25 — .13 = .12 is an estimate of the environmental variance *between families.*

The situation is relatively simple when we deal only with MZ twins, who are genetically identical, or with unrelated children, who have nothing in common genetically. But in order to estimate heritability from any of the other kinship correlations, much more complex formulas are needed which would require much more explanation than is possible in this article. I have presented elsewhere a generalized formula for estimating heritability from any two kinship correlations where one kinship is of a higher degree than the other (Jensen, 1967a). I applied this heritability formula to all the correlations for monozygotic and dizygotic (half their genes in common) twins reported in the literature and found an average heritability of .80 for intelligence test scores. (The correlations from which this heritability estimate was derived were corrected for unreliability.) Environmental differences *between* families account for .12 of the total variance, and differences *within* families account for .08. It is possible to derive an overall heritability coefficient from all the kinship correlations given in Table 2. This composite value of H is .77, which becomes .81 after correction for unreliability (assuming an average test reliability of .95). This represents probably the best single overall estimate of the heritability of measured intelligence that we can make. But, as pointed out previously, this is an average value of H about which there is some dispersion of values, depending on such variables as the particular tests used, the population sampled, and sampling error.

Identical Twins Reared Apart. The conceptually simplest estimate of heritability is, of course, the correlation between identical twins reared apart, since, if their environments are uncorrelated, all they have in common are their genes. The correlation (corrected for unreliability) in this case is the same as the heritability as defined in Equation 3. There have been only three major studies of MZ twins separated early in life and reared apart. All three used individually

administered intelligence tests. The correlation between Stanford-Binet IQs of 19 pairs of MZ twins reared apart in a study by Newman, Freeman, and Holzinger (1937) was .77 (.81 corrected for unreliability). The correlation between 44 pairs of MZ twins reared apart on a composite score based on a vocabulary test and Raven's Progressive Matrices was .77 (.81 corrected) in a study by Shields (1962). The correlation between 53 pairs on the Stanford-Binet was .86 (.91 corrected) in a study by Burt (1966). Twin correlations in the same group for height and for weight were .94 and .88, respectively.

The Burt study is perhaps the most interesting, for four reasons: (*a*) it is based on the largest sample; (*b*) the IQ distribution of the sample had a mean of 97.8 and a standard deviation of 15.3—values very close to those of the general population; (*c*) all the twin pairs were separated at birth or within their first six months of life; and (*d*) most important, the separated twins were spread over the entire range of socioeconomic levels (based on classification in terms of the six socioeconomic categories of the English census), and there was a slight, though nonsignificant, negative correlation between the environmental ratings of the separated twin pairs. When the twin pairs were rated for differences in the cultural conditions of their rearing, these differences correlated .26 with the differences in their IQs. Differences between the material conditions of their homes correlated .16 with IQ differences. (The corresponding correlations for a measure of scholastic attainments were .74 and .37, respectively. The correlation between the twins in scholastic attainments was only .62, indicating a much lower heritability than for IQ.)

Foster Parents versus Natural Parents. Children separated from their true parents shortly after birth and reared in adoptive homes show almost the same degree of correlation with the intelligence of their biological parents as do children who are reared by their own parents. The correlations of children with their foster parents' intelligence range between 0 and .20 and are seldom higher than this even when the adoption agency attempts selective placement (e.g., Honzik, 1957). Parent-child correlations gradually increase from zero at 18 months of age to an asymptotic value close to .50 between ages 5 and 6 (Jones, 1954), and this is true whether the child is reared by his parents or not.

Direct Measurement of the Environment. Another method for getting at the relative contribution of environmental factors to IQ variance is simply by correlating children's IQs with ratings of their environment. This can be legitimately done

only in the case of adopted children and where there is evidence that selective placement by the adoption agencies is negligible. Without these conditions, of course, some of the correlation between the children and their environmental ratings will be due to genetic factors. There are two large-scale studies in the literature which meet these criteria. Also, both studies involved adopting parents who were representative of a broad cross section of the U.S. Caucasian population with respect to education, occupation, and socioeconomic level. It is probably safe to say that not more than five percent of the U.S. Caucasian population falls outside the range of environmental variation represented in the samples in these two studies. The study by Leahy (1935) found an average correlation of .20 between the IQs of adopted children and a number of indices of the "goodness" of their environment, including the IQs and education of both adopting parents, their socioeconomic status, and the cultural amenities in the home. Leahy concluded from this that the environmental ratings accounted for 4 percent (i.e., the square of $r = .20$) of the variance in the adopted children's Stanford-Binet IQs, and that 96 percent of the variance remained to be accounted for by other factors. The main criticisms we can make of this study are, first, that the environmental indices were not sufficiently "fine-grained" to register the subtleties of environmental variation and of the qualities of parent-child relationship that influence intellectual development, and, second, that the study did not make use of the technique of multiple correlation, which would show the total contribution to the variance of all the separate environmental indices simultaneously. A multiple correlation is usually considerably greater than merely the average of all the correlations for the single variables.

A study by Burks (1928) meets both these objections. To the best of my knowledge no study before or since has rated environments in any more detailed and fine-grained manner than did Burks'. Each adoptive home was given 4 to 8 hours of individual investigation. As in Leahy's study, Burks included intelligence measures on the adopting parents as part of the children's environment, an environment which also included such factors as the amount of time the parents spent helping the children with their school work, the amount of time spent reading to the children, and so on. The multiple correlation (corrected for unreliability) between Burks' various environmental ratings and the adopted children's Stanford-Binet IQs was .42. The square of this correlation is .18, which represents the proportion of IQ variance accounted for by Burks' environmental measurements. This value comes very close to the environmental variance estimated in direct heritability analyses based on kinship correlations.

Burks translated her findings into the conclusion that the total effect of environmental factors one standard deviation up or down the environmental scale is only about 6 IQ points. This is an interesting figure, since it is exactly half the 12 point IQ difference found on the average between normal siblings reared together by their own parents. Siblings differ genetically, of course, having only about half their genes in common. If all the siblings in every family were divided into two groups—those above and those below the family average—the IQ distributions of the two groups would appear as shown in Figure 7. Though the average difference is only 12 IQ points, note the implications in the proportions of

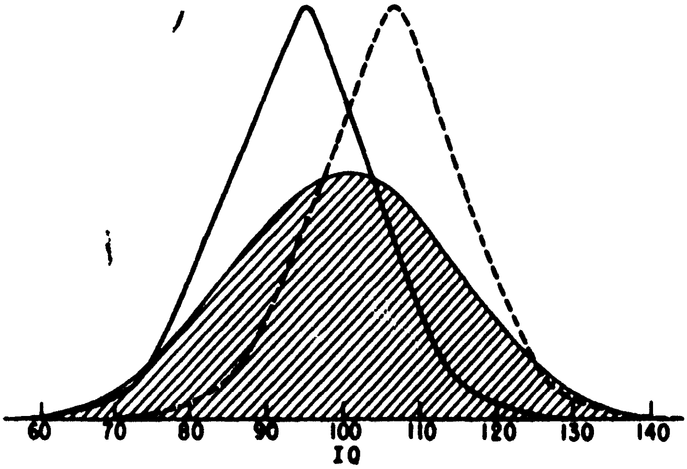


FIGURE 7.
IQ distributions of siblings who are below (solid curve) or above (dashed curve) their family average. The shaded curve is the IQ distribution of randomly selected children.

each group falling into the upper and lower ranges of the IQ scale. It would be most instructive to study the educational and occupational attainments of these two groups, since presumably they should have about the same environmental advantages.

Another part of Burks' study consisted of a perfectly matched control group of parents rearing their own children, for whom parent-child correlations were obtained. Sewall Wright (1931) performed a heritability analysis on these parent-child and IQ-environment correlations and obtained a heritability coefficient of .81.

77-843 O - 72 - 8

Effects of Inbreeding on Intelligence

One of the most impressive lines of evidence for the involvement of genetic factors in intelligence comes from study of the effects of inbreeding, that is, the mating of relatives. In the case of polygenic characteristics the direction of the effect of inbreeding is predictable from purely genetic considerations. All individuals carry in their chromosomes a number of mutant or defective genes. These genes are almost always recessive, so they have no effect on the phenotype unless by rare chance they match up with another mutant gene at the same locus on a homologous chromosome; in other words, the recessive mutant gene at a given locus must be inherited from both the father and mother in order to affect the phenotype. Since such mutants are usually defective, they do not enhance the phenotypic expression of the characteristic but usually degrade it. And for polygenic characteristics we would expect such mutants to lower the metric value of the characteristics by graded amounts, depending upon the number of paired mutant recessives. If the parents are genetically related, there is a greatly increased probability that the mutant recessives at given loci will be paired in the offspring. The situation is illustrated in Figure 8, which depicts in a simplified way a pair of homologous chromosomes inherited by an individual from a mother (M) and father (F) who are related (Pair A) and a pair of chromosomes inherited from unrelated parents (Pair B). The blackened spaces represent recessive genes. Although both pairs contain equal numbers of recessives, more of them are at the same loci in Pair A than in Pair B. Only the paired genes degrade the characteristics' phenotypic value.

A most valuable study of this genetic phenomenon with respect to intelligence was carried out in Japan after World War II by Schull and Neel (1965). The study illustrates how strictly sociological factors, such as mate selection, can have extremely important genetic consequences. In Japan approximately five percent of all marriages are between cousins. Schull and Neel studied the offspring of marriages of first cousins, first cousins once removed, and second cousins. The parents were statistically matched with a control group of unrelated parents for age and socioeconomic factors. Children from the cousin marriages and the control children from unrelated parents (total $N = 2{,}111$) were given the Japanese version of the Wechsler Intelligence Scale for Children (WISC). The degree of consanguinity represented by the cousin marriages in this study had the effect of depressing WISC IQs by an average of 7.4 percent, making the mean of the inbred group nearly 8 IQ points lower than the mean of the control group. Assuming normal distributions of IQ, the effect is shown in

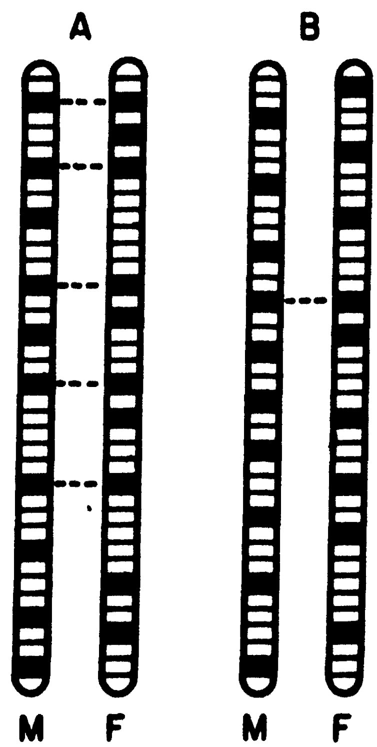


FIGURE 8.

Simplified schema of chromosomes, illustrating the pairing of recessive (mutant) genes (black spaces) in homologous chromosomes from mother (M) and father (F). Pair A has five pairs of recessives in the same loci on the chromosome, Pair B has only one such pair.

Figure 9, and illustrates the point that the most drastic consequences of group mean differences are to be seen in the tails of the distributions. In the same study a similar depressing effect was found for other polygenic characteristics such as several anthropometric and dental variables.

The mating of relatives closer than cousins can produce a markedly greater reduction in offspring's IQs. Lindzey (1967) has reported that almost half of a group of children born to so-called nuclear incest matings (brother-sister or father-daughter) could not be placed for adoption because of mental retardation and other severe defects which had a relatively low incidence among the offspring of unrelated parents who were matched with the incestuous parents in

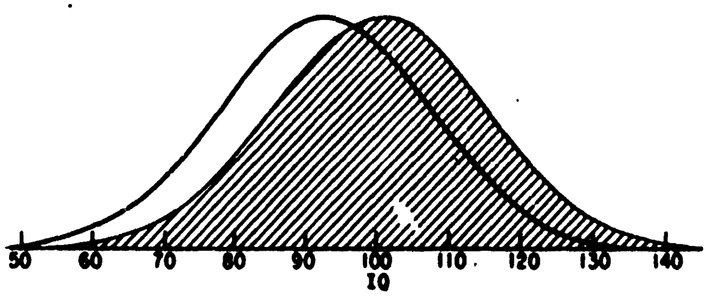


FIGURE 9.

The average effect of inbreeding to the degree of 1st, 1 1/2, and 2nd cousin matings on the IQ distribution of offspring (heavy line). Shaded curve is the IQ distribution of the offspring of nonconsanguinous matings. (After Schull & Neel, 1965.)

intelligence, socioeconomic status, age, weight, and stature. In any geographically confined population where social or legal regulations on mating are lax, where individuals' paternity is often dubious, and where the proportion of half-siblings within the same age groups is high, we would expect more inadvertent inbreeding, with its unfavorable genetic consequences, than in a population in which these conditions exist to a lesser degree.

Heritability of Special Mental Abilities. When the general factor, or *g*, is removed from a variety of mental tests, the remaining variance is attributable to a number of so-called "group factors" or "special abilities." The tests of special abilities that have been studied most thoroughly with respect to their heritability are Thurstone's Primary Mental Abilities: Verbal, Space, Number, Word Fluency, Memory, and Perceptual Speed. Vandenberg (1967) has reviewed the heritability studies of these tests and reports that the *H* values range from near zero to about .75, with most values of *H* between .50 and .70. Vandenberg devised a method for estimating the genetic components of these special abilities which are completely independent of *g*. He concluded that at least four of the Primary Mental Abilities (Number, Verbal, Space, and Word Fluency) independently have significant hereditary components.

There have been few studies of the heritability of noncognitive skills, but a study by McNemar (see Bilodeau, 1966, Ch. 3) of motor skill learning indicates that heritabilities in this sphere may be even higher than for intelligence. The

motor skill learning was measured with a pursuit-rotor, a tracking task in which the subject must learn to keep a stylus on a metal disc about the size of a nickel rotating through a circumference of about 36 inches at 60 rpm. The percentage of time "on target" during the course of practice yields a learning measure of high reliability, showing marked individual differences both in rate of acquisition and final asymptote of this perceptual-motor skill. Identical twins correlated .95 and fraternal twins .51 on pursuit-rotor learning, yielding a heritability coefficient of .88, which is very close to the heritability of physical stature.

Heritability of Scholastic Achievement. The heritability of measures of scholastic achievement is much less, on the average, than the heritability of intelligence. In reviewing all the twin studies in the literature containing relevant data, I concluded that individual differences in scholastic performance are determined less than half as much by heredity as are differences in intelligence (Jensen, 1967a).[8] The analysis of all the twin studies on a variety of scholastic measures gives an average H of .40. The environmental variance of 60 percent can be partitioned into variance due to environmental differences *between* families, which is 54 percent, and differences *within* families of 6 percent. But it should also be noted that the heritability estimates for scholastic achievement vary over a much wider range than do H values for intelligence. In general, H for scholastic achievement increases as we go from the primary grades up to high school and it is somewhat lower for relatively simple forms of learning (e.g., spelling and arithmetic compu-

[8] After this article went to press I received a personal communication from Professor Lloyd G. Humphreys who pointed out some arguments that indicate I may have underestimated the heritability of scholastic achievement and that its heritability may actually be considerably closer to the heritability of intelligence. The argument involves two main points: (1) the fact that some of the achievement tests that entered into the average estimate of heritability are tests of specific achievements, rather than omnibus achievement tests, and therefore would correspond more to the separate subscales of the usual intelligence tests, which are known to have somewhat lower heritabilities than the composite scores; and (2) scores on some of the achievement tests are age-related, so that fraternal twin correlations, in relation to other kinship correlations, are unduly inflated by common factor of age. When age is partialled out of the MZ and DZ twin correlations, the estimate of heritability based on MZ and DZ twin comparisons is increased. However, an omnibus achievement test (Stanford Achievement) yielding an overall Educational Age score had a heritability of only .48 (as compared with .63 for Stanford-Binet IQ and .70 for Otis IQ based on the same set of MZ and DZ twins), with age partialled out of the twin correlations (Newman, Freeman, and Holzinger, 1937, p. 97). Rank in high school graduating class, which is an overall index of scholastic performance and is little affected by age yields heritability coefficients below .40 in a nationwide sample (Nichols & Bilbro, 1966). The issue clearly needs further study, but the best conclusion that can be drawn from the existing evidence, I believe, still is that the heritability of scholastic achievement is less than for intelligence, but the amount of the difference cannot be precisely estimated at present.

tation) than for more complex learning (e.g., reading comprehension and arithmetic problem solving). Yet large-sample twin data from the National Merit Scholarship Corporation show that the *between families* environmental component accounts for about 60 percent of the variance in students' rank in their high school graduating class. This must mean that there are strong family influences which cause children to conform to some academic standard set by the family and which reduce variance in scholastic performance among siblings reared in the same family. Unrelated children reared together are also much more alike in school performance than in intelligence. The common finding of a negative correlation between children's IQ and the amount of time parents report spending in helping their children with school work is further evidence that considerable family pressures are exerted to equalize the scholastic performance of siblings. This pressure to conform to a family standard shows up most conspicuously in the small *within families* environmental variance component on those school subjects which are most susceptible to improvement by extra coaching, such as spelling and arithmetic computation.

The fact that scholastic achievement is considerably less heritable than intelligence also means that many other traits, habits, attitudes, and values enter into a child's performance in school besides just his intelligence, and these non-cognitive factors are largely environmentally determined, mainly through influences within the child's family. This means there is potentially much more we can do to improve school performance through environmental means than we can do to change intelligence per se. Thus it seems likely that if compensatory education programs are to have a beneficial effect on achievement, it will be through their influence on motivation, values, and other environmentally conditioned habits that play an important part in scholastic performance, rather than through any marked direct influence on intelligence per se. The proper evaluation of such programs should therefore be sought in their effects on actual scholastic performance rather than in how much they raise the child's IQ.

How the Environment Works

Environment as a Threshold

All the reports I have found of especially large upward shifts in IQ which are explicitly associated with environmental factors have involved young children, usually under six years of age, whose initial social environment was deplorable to a greater extreme than can be found among any children who are free to inter-

act with other persons or to run about out-of-doors. There can be no doubt that moving children from an extremely deprived environment to good average environmental circumstances can boost the IQ some 20 to 30 points and in certain extreme rare cases as much as 60 or 70 points. On the other hand, children reared in rather average circumstances do not show an appreciable IQ gain as a result of being placed in a more culturally enriched environment. While there are reports of groups of children going from below average up to average IQs as a result of environmental enrichment, I have found no report of a group of children being given permanently superior IQs by means of environmental manipulations. In brief, it is doubtful that psychologists have found consistent evidence for any social environmental influences short of extreme environmental isolation which have a marked systematic effect on intelligence. This suggests that the influence of the quality of the environment on intellectual development is not a linear function. Below a certain threshold of environmental adequacy, deprivation can have a markedly depressing effect on intelligence. But above this threshold, environmental variations cause relatively small differences in intelligence. The fact that the vast majority of the populations sampled in studies of the heritability of intelligence are above this threshold level of environmental adequacy accounts for the high values of the heritability estimates and the relatively small proportion of IQ variance attributable to environmental influences.

The environment with respect to intelligence is thus analogous to nutrition with respect to stature. If there are great nutritional lacks, growth is stunted, but above a certain level of nutritional adequacy, including minimal daily requirements of minerals, vitamins, and proteins, even great variations in eating habits will have negligible effects on persons' stature, and under such conditions most of the differences in stature among individuals will be due to heredity.

When I speak of subthreshold environmental deprivation, I do not refer to a mere lack of middle-class amenities. I refer to the extreme sensory and motor restrictions in environments such as those described by Skeels and Dye (1939) and Davis (1947), in which the subjects had little sensory stimulation of any kind and little contact with adults. These cases of extreme social isolation early in life showed great deficiencies in IQ. But removal from social deprivation to a good, average social environment resulted in large gains in IQ. The Skeels and Dye orphanage children gained in IQ from an average of 64 at 19 months of age to 96 at age 6 as a result of being given social stimulation and placement in good homes between 2 and 3 years of age. When these children were followed up as adults, they were found to be average citizens in their communities, and their own

children had an average IQ of 105 and were doing satisfactorily in school. A far more extreme case was that of Isabel, a child who was confined and reared in an attic up to the age of six by a deaf-mute mother, and who had an IQ of about 30 at age 6. When Isabel was put into a good environment at that age, her IQ became normal by age 8 and she was able to perform as an average student throughout school (Davis, 1947). Extreme environmental deprivation thus need not permanently result in below average intelligence.

These observations are consistent with studies of the effects of extreme sensory deprivation on primates. Monkeys raised from birth under conditions of total social isolation, for example, show no indication when compared with normally raised controls, of any permanent impairment of ability for complex discrimination learning, delayed response learning, or learning set formation, although the isolated monkeys show severe social impairment in their relationships to normally reared monkeys (Harlow & Griffin, 1965).

Thoughtful scrutiny of all these studies of extreme environmental deprivation leads to two observations which are rarely made by psychologists who cite the studies as illustrative explanations of the low IQs and poor scholastic performance of the many children called culturally disadvantaged. In the first place, typical culturally disadvantaged children are not reared in anything like the degree of sensory and motor deprivation that characterizes, say, the children of the Skeels study. Secondly, the IQs of severely deprived children are markedly depressed even at a very early age, and when they are later exposed to normal environmental stimulation, their IQs rise rapidly, markedly, and permanently. Children called culturally disadvantaged, on the other hand, generally show no early deficit and are usually average and sometimes precocious on perceptual-motor tests administered before two years of age. The orphanage children described in Skeels' study are in striking contrast to typical culturally disadvantaged children of the same age. Also, culturally disadvantaged children usually show a slight initial gain in IQ after their first few months of exposure to the environmental enrichment afforded by school attendance, but, unlike Skeels' orphans, they soon lose this gain, and in a sizeable proportion of children the initial IQ gain is followed by a gradual decline in IQ throughout the subsequent years of schooling. We do not know how much of this decline is related to environmental or hereditary factors. We do know that with increasing age children's IQs increasingly resemble their parents' rank order in intelligence whether they are reared by them or not, and therefore with increasing age we should expect greater and more reliable differentiation among children's IQs as they gravitate toward their genotypic values

(Honzik, 1957). Of course, the gravitating effect is compounded by the fact that less intelligent parents are also less apt to provide the environmental conditions conducive to intellectual development in the important period between ages 3 and 7, during which children normally gain increasing verbal control over their environment and their own behavior. (I have described some of these environmental factors in detail elsewhere [Jensen, 1968e].)

Heber, Dever, and Conry (1968) have obtained data which illustrate this phenomenon of children's gravitation toward the parental IQ with increasing age. They studied the families of 88 low economic class Negro mothers residing in Milwaukee in a set of contiguous slum census tracts, an area which yields the highest known prevalence of identified retardation in the city's schools. Although these tracts contribute about 5 percent of the schools' population, they account for about one-third of the school children classed as mentally retarded (IQ below 75). The sample of 88 mothers was selected by taking 88 consecutive births in these tracts where the mother already had at least one child of age six. The 88 mothers had a total of 586 children, excluding their newborns. The percentage of mothers with IQs of 80 or above was 54.6; 45.4 percent were below IQ 80. The IQs of the children of these two groups of mothers were plotted as a function of the children's age. The results are shown in Figure 10. Note that only the children whose mothers' IQs are below 80 show a systematic decline in IQ as well as a short-lived spurt of several points at the age of entrance into school. At six years of age and older, 80.8 percent of the children with IQs below 80 were those whose mothers had IQs below 80.

It is far from certain or even likely that all such decline in IQ is due to environmental influences rather than to genetic factors involved in the growth rate of intelligence. Consistent with this interpretation is the fact that the heritability of intelligence measures increases with age. We should expect just the opposite if environmental factors alone were responsible for the increasing IQ deficit of markedly below average groups. A study by Wheeler (1942) suggests that although IQ may be raised at all age levels by improving the environment, such improvements do not counteract the decline in the IQ of certain below-average groups. In 1940 Wheeler tested over 3000 Tennessee mountain children between the ages of 6 and 16 and compared their IQs with children in the same age range who had been given the same tests in 1930, when the average IQ and standard of living in this area would characterize the majority of the inhabitants as "culturally deprived." During the intervening 10 years state and federal intervention in this area brought about great improvements in economic conditions, standards

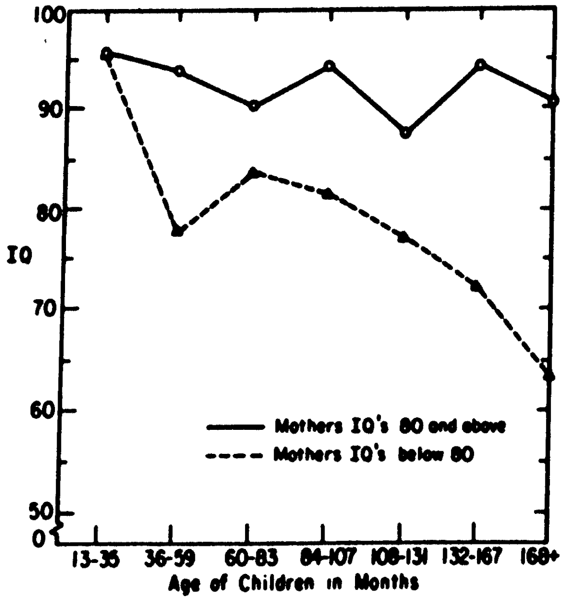


FIGURE 10.
Mean IQs of 586 children of 88 mothers as a function of age of children. (Heber, Dever, & Conry, 1968.)

of health care, and educational and cultural opportunities, and during the same period the average IQ for the region increased 10 points, from 82 to 92. But the decline in IQ from age 6 to age 16 was about the same in 1940 (from 103 to 80) as in 1930 (from 95 to 74).

Reaction Range. Geneticists refer to the concept of reaction range (RR) in discussing the fact that similar genotypes may result in quite different phenotypes depending on the favorableness of the environment for the development of the characteristic in question. Of further interest to geneticists is the fact that different genotypes may have quite different reaction ranges; some genotypes may be much more buffered against environmental influences than others. Different genetic strains can be unequal in their susceptibility to the same range of environmental variation, and when this is the case, the strains will show dissimilar heritabilities on the trait in question, the dissimilarity being accentuated by increasing environmental variation. Both of these aspects of the reaction range concept are illustrated hypothetically with respect to IQ in Figure 11.

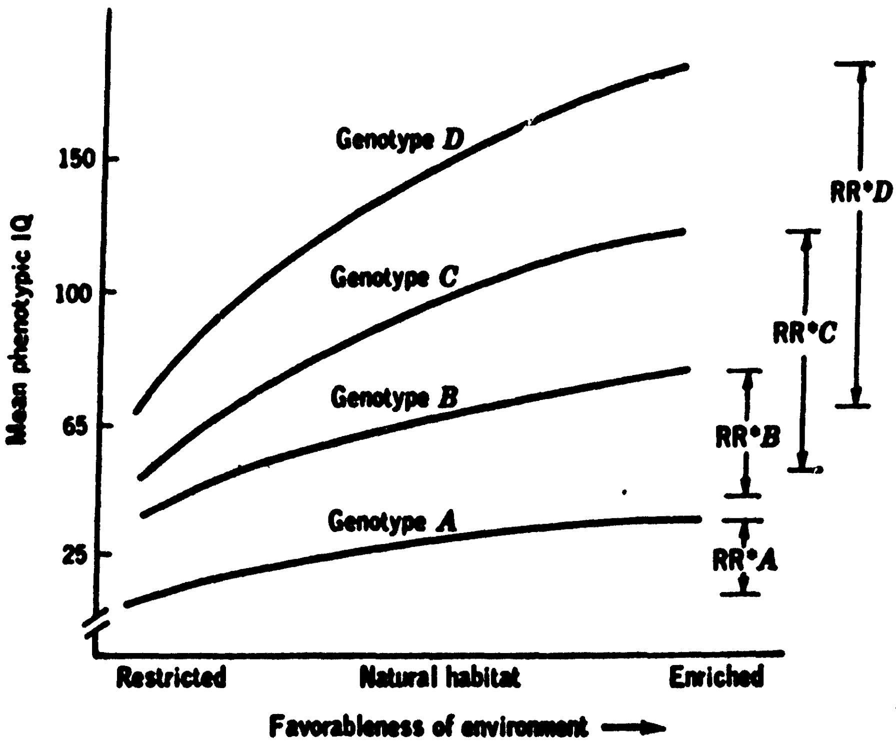


FIGURE 11.

Scheme of the reaction range concept for four hypothetical genotypes. RR denotes the presumed reaction range for phenotypic IQ. Note: Large deviations from the "natural habitat" have a low probability of occurrence. (From Gottesman, 1963.)

The above discussion should serve to counter a common misunderstanding about quantitative estimates of heritability. It is sometimes forgotten that such estimates actually represent *average* values in the population that has been sampled and they do not necessarily apply either to differences *within* various subpopulations or to differences *between* subpopulations. In a population in which an overall *H* estimate is, say, .80, we may find a certain group for which *H* is only .70 and another group for which *H* is .90. All the major heritability studies reported in the literature are based on samples of white European and North American populations, and our knowledge of the heritability of intelligence in different racial and cultural groups within these populations is nil. For example,

no adequate heritability studies have been based on samples of the Negro population of the United States. Since some genetic strains may be more buffered from environmental influences than others, it is not sufficient merely to equate the environments of various subgroups in the population to infer equal heritability of some characteristic in all of them. The question of whether heritability estimates can contribute anything to our understanding of the relative importance of genetic and environmental factors in accounting for average phenotypic differences between racial groups (or any other socially identifiable groups) is too complex to be considered here. I have discussed this problem in detail elsewhere and concluded that heritability estimates could be of value in testing certain specific hypotheses in this area of inquiry, provided certain conditions were met and certain other crucial items of information were also available (Jensen, 1968c).

Before continuing discussion of environmental factors we must guard against one other misunderstanding about heritability that sometimes creeps in at this point. This is the notion that because so many different environmental factors and all their interactions influence the development of intelligence, by the time the child is old enough to be tested, these influences must totally bury or obscure all traces of genetic factors—the genotype must lie hidden and inaccessible under the heavy overlay of environmental influences. If this were so, of course, the obtained values of H would be very close to zero. But the fact that values of H for intelligence are usually quite high (in the region of .70 to .90) means that current intelligence tests can, so to speak, "read through" the environmental "overlay."

Physical versus Social Environment

The value $1 - H$, which for IQ generally amounts to about .20, can be called E, the proportion of variance due to nongenetic factors. There has been a pronounced tendency to think of E as being wholly associated with individuals' social and interpersonal environment, child rearing practices, and differences in educational and cultural opportunities afforded by socioeconomic status. It is certain, however, that these sociological factors are not responsible for the whole of E and it is not improbable that they contribute only a minor portion of the E variance in the bulk of our population. Certain physical and biological environmental factors may be at least as important as the social factors in determining individual differences in intelligence. If this is true, advances in medicine, nutrition, prenatal care, and obstetrics may contribute as much or more to improving intelligence as will manipulation of the social environment.

Prenatal Environment of Twins. A little known fact about twins is that they average some 4 to 7 points lower in IQ than singletons (Vandenberg, 1968). The difference also shows up in scholastic achievement, as shown in the distribution of reading scores of twin and singleton girls in Sweden (Figure 12).

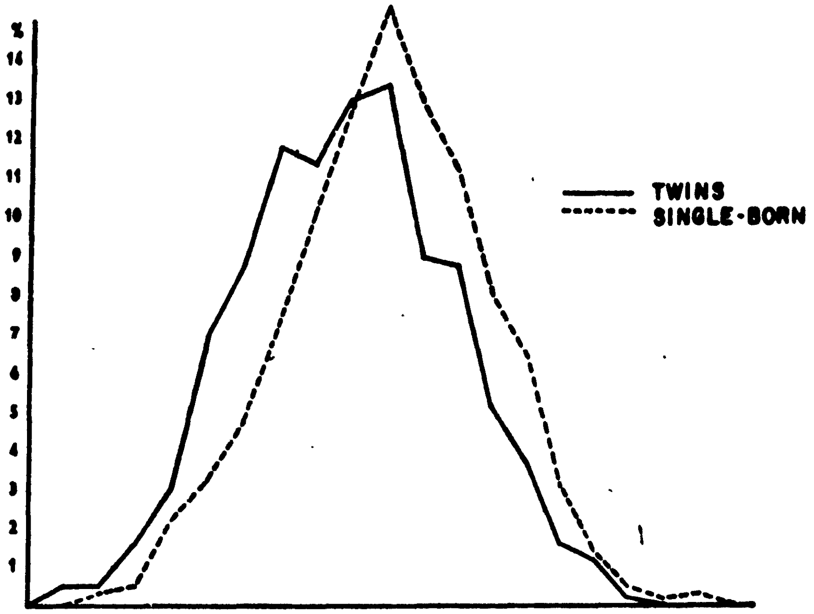


FIGURE 12.
Distribution of reading scores of twins and single children (all girls). (Husén, 1960.)

If this phenomenon were due entirely to differences between twins and singletons in the amount of individual attention they receive from their parents, one might expect the twin-singleton difference to be related to the family's socioeconomic status. But there seems to be no systematic relationship of this kind. The largest study of the question, summarized in Figure 13, shows about the same average amount of twin-singleton IQ disparity over a wide range of socioeconomic groups.

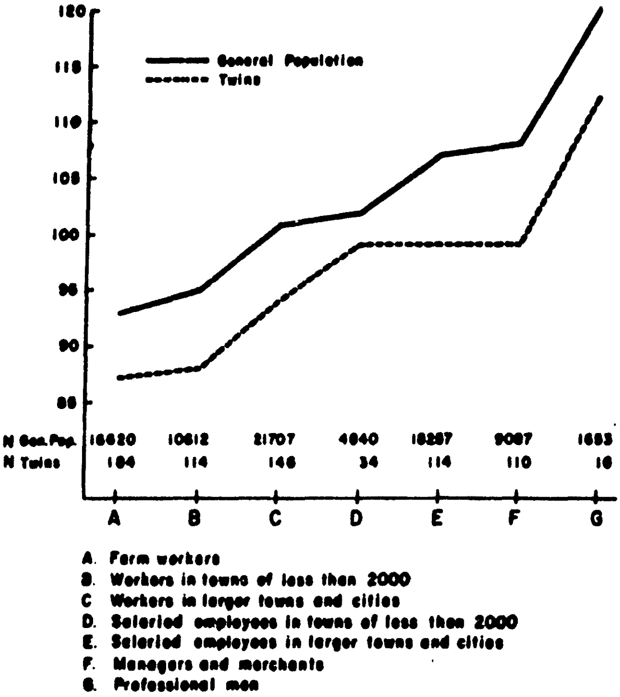


FIGURE 13.

Distribution of IQs by occupation of father, for twins and singletons. (Zazzo, 1960.)

Three other lines of evidence place the locus of this effect in the prenatal environment. Monozygotic twins are slightly lower in IQ than dizygotic twins (Stott, 1960, p. 98), a fact which is consistent with the finding that MZ twins have a higher mortality rate and greater disparity in birth weights than DZ twins, suggesting that MZ twins enjoy less equal and less optimal intrauterine conditions than DZ twins or singletons. Inequalities in both intrauterine space and fetal nutrition probably account for this. Also, boy twins are significantly lower in IQ than girl twins, which conforms to the well known greater vulnerability of male infants to prenatal impairment (Stott, 1960). Finally, the birth weight of infants, when matched for gestational age, is slightly but significantly correlated with later IQ, and the effect is independent of sociocultural factors (Churchill, Neff,

& Caldwell, 1966). In pairs of identical twins, the twin with the lower birth weight usually has the lower IQ (by 5 to 7 points on the average) at school age. This is true both in white and in Negro twins. The birth-weight differences are reflected in all 11 subtests of the Wechsler Intelligence Scale for Children and are slightly greater on the Performance than on the Verbal tests (Willerman & Churchill, 1967). The investigators interpret these findings as suggesting that nutrient supplies may be inadequate for proper body and brain development in twin pregnancies, and that the unequal sharing of nutrients and space stunts one twin more than its mate.

Thus, much of the average difference between MZ twins, whether reared together or reared apart, seems to be due to prenatal environmental factors. The real importance of these findings, of course, lies in their implications for the possible role of prenatal environment in the development of all children. It is not unlikely that there are individual maternal differences in the adequacy of the prenatal environment. If intrauterine conditions can cause several points of IQ difference between twins, it is not hard to imagine that individual differences in prenatal environments could also cause IQ differences in single born children and might therefore account for a substantial proportion of the total environmental variance in IQ.

Abdominal Decompression. There is now evidence that certain manipulations of the intrauterine environment can affect the infant's behavioral development for many months after birth. A technique known as abdominal decompression was invented by a professor of obstetrics (Heyns, 1963), originally for the purpose of making women experience less discomfort in the latter months of their pregnancy and also to facilitate labor and delivery. For about an hour a day during the last three or four months of pregnancy, the woman is placed in a device that creates a partial vacuum around her abdomen, which greatly reduces the intrauterine pressure. The device is used during labor up to the moment of delivery. Heyns has applied this device to more than 400 women. Their infants, as compared with control groups who have not received this treatment, show more rapid development in their first two years and manifest an overall superiority in tests of perceptual-motor development. They sit up earlier, walk earlier, talk earlier, and appear generally more precocious than their own siblings or other children whose mothers were not so treated. At two years of age the children in Heyns' experiment had DQs (developmental quotients) some 30 points higher than the control children (in the general population the mean DQ is 100, with

a standard deviation of 15). Heyns explains the effects of maternal abdominal decompression on the child's early development in terms of the reduction of intrauterine pressure, which results in a more optimal blood supply to the fetus and also lessens the chances of brain damage during labor. (The intrauterine pressure on the infant's head is reduced from about 22 pounds to 8 pounds.) Results on children's later IQs have not been published, but correspondence with Professor Heyns and verbal reports from visitors to his laboratory inform me that there is no evidence that the IQ of these children is appreciably higher beyond age 6 than that of control groups. If this observation is confirmed by the proper methods, it should not be too surprising in view of the negligible correlations normally found between DQs and later IQs. But since abdominal decompression results in infant precocity, one may wonder to what extent differences in intrauterine pressure are responsible for normal individual and group differences in infant precocity. Negro infants, for example, are more precocious in development (as measured on the Bayley Scales) in their first year or two than Caucasian infants (Bayley, 1965a). Infant precocity would seem to be associated with more optimal intrauterine and perinatal conditions. This conjecture is consistent with the finding that infants whose prenatal and perinatal histories would make them suspect of some degree of brain damage show lower DQs on the Bayley Scales than normal infants (Honzik, 1962). Writers who place great emphasis on the hypothesis of inadequate prenatal care and complications of pregnancy to account for the lower average IQ of Negroes (e.g., Bronfenbrenner, 1967) are also obliged to explain why these unfavorable factors do not also depress the DQ below average in Negro infants, as do such factors as brain damage and prenatal and infant malnutrition (Cravioto, 1966). Since all such environmental factors should lower the heritability of intelligence in any segment of the population in which they are hypothesized to play an especially significant role, one way to test the hypothesis would be to compare the heritability of intelligence in that segment of the population for which extra environmental factors are hypothesized with the heritability in other groups for whom environmental factors are supposedly less accountable for IQ variance.

A Continuum of Reproductive Casualty. A host of conditions associated with reproduction which are known to differ greatly across socioeconomic levels have been hypothesized as causal factors in average intellectual differences. There is no doubt about the fact of the greater prevalence in poverty areas of conditions unfavorable to optimal pregnancy and safe delivery. The question that remains

unanswered is the amount of IQ variance associated with these conditions predisposing to reproductive casualty. The disadvantageous factors most highly associated with social conditions are: pregnancies at early ages, teenage deliveries, pregnancies in close succession, a large number of pregnancies, and pregnancies that occur late in the woman's reproductive life (Graves, Freeman, & Thompson, 1968). These conditions are related to low birth weight, prematurity, increased infant mortality, prolonged labor, toxemia, anemia, malformations, and mental deficiency in the offspring. Since all of these factors have a higher incidence in low socioeconomic groups and in certain ethnic groups (Negroes, American Indians, and Mexican-Americans) in the United States, they probably account for some proportion of the group differences in IQ and scholastic performance, but just how much of the true differences they may account for no one really knows at present. It is interesting that Jewish immigrants, whose offspring are usually found to have a higher mean IQ than the general population, show fewer disadvantageous reproductive conditions and have the lowest infant mortality rates of all ethnic groups, even when matched with other immigrant and native born groups on general environmental conditions (Graves et al., 1968).

Although disadvantageous reproductive factors occur differentially in different segments of the population, it is not at all certain how much they are responsible for the IQ differences between social classes and races. It is reported by the National Institute of Neurological Diseases and Blindness, for example, that when all cases of mental retardation that can be reasonably explained in terms of known complications of pregnancy and delivery, brain damage, or major gene and chromosomal defects are accounted for, there still remain 75 to 80 percent of the cases who show no such specific causes and presumably represent just the lower end of the normal polygenic distribution of intelligence (Research Profile No. 11, 1965). Buck (1968) has argued that it still remains to be proven that a degree of neurological damage is bound to occur among the survivors of all situations which carry a high risk of perinatal mortality and that a high or even a known proportion of mental retardation can be ascribed to the non-lethal grades of reproductive difficulty. A large study reported by Buck (1968) indicates that the most common reproductive difficulties when occurring singly have no significant effect on children's intellectual status after age 5, with the one exception of pre-eclamptic toxemia of pregnancy, which caused some cognitive impairment. Most of the complications of pregnancy, it seems, must occur multiply to impair intellectual ability. It is as if the nervous system is sufficiently homeostatic to withstand certain unfavorable conditions if they occur singly.

77-048 O - 73 - 10

Prematurity. The literature on the relationship of premature birth to the child's IQ is confusing and conflicting. Guilford (1967), in his recent book on *The Nature of Intelligence*, for example, concluded, as did Stoddard (1943), that prematurity has no effect on intelligence. Stott (1966), on the other hand, presents impressive evidence of very significant IQ decrements associated with prematurity. Probably the most thorough review of the subject I have found, by Kushlick (1966), helps to resolve these conflicting opinions. There is little question that prematurity has the strongest known relation to brain dysfunction of any reproductive factor, and many of the complications of pregnancy are strongly associated with the production of premature children. The crucial factor in prematurity, however, is not prematurity per se, but low birth-weight. Birth-weight apparently acts as a threshold variable with respect to intellectual impairment. All studies of birth-weight agree in showing that the incidence of babies weighing less than 5-1/2 lbs. increases from higher to lower social classes. But only about 1 percent of the total variance of birth-weight is accounted for by socioeconomic variables. Race (Negro versus white) has an effect on birth-weight independently of socioeconomic variables. Negro babies mature at a lower birth-weight than white babies (Naylor & Myrianthopoulos, 1967). If prematurity is defined as a condition in which birth-weight is under 5-1/2 lbs., the observed relationship between prematurity and depression of the IQ is due to the common factor of low social class. Kushlick (1966, p. 143) concludes that it is only among children having birth-weights under 3 lbs. that the mean IQ is lowered, independently of social class, and more in boys than in girls. The incidence of extreme subnormality is higher for children with birth-weights under 3 or 4 lbs. But when one does not count these extreme cases (IQs below 50), the effects of prematurity or low birth-weight—even as low as 3 lbs.—have a very weak relationship to children's IQs by the time they are of school age. The association between very low birth-weight and extreme mental subnormality raises the question of whether the low birth-weight causes the abnormality or whether the abnormality arises independently and causes the low birth-weight.

Prematurity and low birth-weight have a markedly higher incidence among Negroes than among whites. That birth-weight differences per se are not a predominant factor in Negro-white IQ differences, however, is suggested by the findings of a study which compared Negro and white premature children matched for birth-weight. The Negro children in all weight groups performed significantly less well on mental tests at 3 and 5 years of age than the white children of comparable birth-weight (Hardy, 1965, p. 51).

Genetic Predisposition to Prenatal Impairment. Dennis Stott (1960, 1966), a British psychologist, has adduced considerable evidence for the theory that impairments of the central nervous system occurring prenatally as a result of various stresses in pregnancy may not be the *direct* result of adverse intrauterine factors but may result *indirectly* from genetically determined mechanisms which are triggered by prenatal stress of one form or another.

Why should there exist a genetic mechanism predisposing to congenital impairments? Would not such genes, if they had ever existed, have been eliminated long ago through natural selection? It can be argued from considerable evidence in lower species of mammals observable by zoologists today that such a genetic mechanism may have had survival value for primitive man, but that the conditions of our present industrial society and advances in medical care have diminished the biological advantage of this mechanism for survival of the human species. The argument is that, because of the need to control population, there is a genetic provision within all species for multiple impairments, which are normally only potentialities, that can be triggered off by prenatal stress associated with high population density, such as malnutrition, fatigue from overexertion, emotional distress, infections, and the like. The resulting congenital impairment would tend to cut down the infant population, thereby relieving the pressure of population without appreciably reducing the functioning and efficiency of the young adults in the population. Stott (1966) has presented direct evidence of an association between stresses in the mother during pregnancy and later behavioral abnormalities and learning problems of the child in school. The imperfect correlation between such prenatal stress factors and signs of congenital impairment suggests that there are individual differences in genetic predisposition to prenatal impairment. The hypothesis warrants further investigation. The prenatal environment could be a much more important source of later IQ variance for some children than for others.

Mother-Child Rh Incompatibility. The *Rh* blood factor can involve possible brain damaging effects in a small proportion of pregnancies where the fetus is *Rh*-positive and the mother is *Rh*-negative. (*Rh*-negative has a frequency of 15 percent in the white and 7 percent in the Negro population.) The mother-child *Rh* incompatibility produces significant physical ill effects in only a fraction of cases and increases in importance in pregnancies beyond the first. The general finding of slightly lower IQs in second and later born children could be related to *Rh* incompatibility or to similar, but as yet undiscovered, mother-child biological incompatibilities. This is clearly an area greatly in need of pioneering research.

Nutrition. Since the human brain attains 70 percent of its maximum adult weight in the first year after birth, it should not be surprising that prenatal and infant nutrition can have significant effects on brain development. Brain growth is largely a process of protein synthesis. During the prenatal period and the first postnatal year the brain normally absorbs large amounts of protein nutrients and grows at the average rate of 1 to 2 milligrams per minute (Stoch & Smythe, 1963; Cravioto, 1966).

Severe undernutrition before two or three years of age, especially a lack of proteins and the vitamins and minerals essential for their anabolism, results in lowered intelligence. Stoch and Smythe (1963) found, for example, that extremely malnourished South African colored children were some 20 points lower in IQ than children of similar parents who had not suffered from malnutrition. The difference between the undernourished group and the control group in DQ and IQ over the age range from 1 year to 8 years was practically constant. If undernutrition takes a toll, it takes it early, as shown by the lower DQs at 1 year and the absence of any increase in the decrement at later ages. Undernutrition occurring for the first time in older children seems to have no permanent effect. Severely malnourished war prisoners, for example, function intellectually at their expected level when they are returned to normal living conditions. The study by Stoch and Smythe, like several others (Cravioto, 1966; Scrimshaw, 1968), also revealed that the undernourished children had smaller stature and head circumference than the control children. Although there is no correlation between intelligence and head circumference in normally nourished children, there is a positive correlation between these factors in groups whose numbers suffer varying degrees of undernutrition early in life. Undernutrition also increases the correlation between intelligence and physical stature. These correlations provide us with an index which could aid the study of IQ deficits due to undernutrition in selected populations.

One of the most interesting and pronounced psychological effects of undernutrition is retardation in the development of cross-modal transfer or intersensory integration, which was earlier described as characterizing the essence of *g* (Scrimshaw, 1968).

The earlier the age at which nutritional therapy is instituted, of course, the more beneficial are its effects. But even as late as 2 years of age, a gain of as much as 18 IQ points was produced by nutritional improvements in a group of extremely undernourished children. After 4 years of age, however, nutritional therapy effected no significant change in IQ (Cravioto, 1966, p. 82).

These studies were done in countries where extreme undernutrition is not uncommon. Such gross nutritional deprivation is rare in the United States. But there is at least one study which shows that some undetermined proportion of the urban population in the United States might benefit substantially with respect to intellectual development by improved nutrition. In New York City, women of low socioeconomic status were given vitamin and mineral supplements during pregnancy. These women gave birth to children who, at four years of age, averaged 8 points higher in IQ than a control group of children whose mothers had been given placebos during pregnancy (Harrell, Woodyard, & Gates, 1955). Vitamin and mineral supplements are, of course, beneficial in this way only when they remedy an existing deficiency.

Birth Order. Order of birth contributes a significant proportion of the variance in mental ability. On the average, first-born children are superior in almost every way, mentally and physically. This is the consistent finding of many studies (Altus, 1966) but as yet the phenomenon remains unexplained. (Rimland [1964, pp. 140-148] has put forth some interesting hypotheses to explain the superiority of the first-born.) Since the first-born effect is found throughout all social classes in many countries and has shown up in studies over the past 80 years (it was first noted by Galton), it is probably a biological rather than a social-psychological phenomenon. It is almost certainly not a genetic effect. (It would tend to make for slightly lower estimates of heritability based on sibling comparisons.) It is one of the sources of environmental variance in ability without any significant postnatal environmental correlates. No way is known for giving later-born children the same advantage. The disadvantage of being later-born, however, is very slight and shows up conspicuously only in the extreme upper tail of the distribution of achievements. For example, there is a disproportionate number of first-born individuals whose biographies appear in *Who's Who* and in the *Encyclopedia Britannica*.

Social Class Differences in Intelligence

Social class (or socioeconomic status [SES]) should be considered as a factor separate from race. I have tried to avoid using the terms *social class* and *race* synonymously or interchangeably in my writings, and I observe this distinction here. Social classes completely cut across all racial groups. But different racial groups are disproportionately represented in different SES categories. Social class differences refer to a socioeconomic continuum *within* racial groups.

It is well known that children's IQs, by school age, are correlated with the socioeconomic status of their parents. This is a world-wide phenomenon and has an extensive research literature going back 70 years. Half of all the correlations between SES and children's IQs reported in the literature fall between .25 and .50, with most falling in the region of .35 to .40. When school children are grouped by SES, the mean IQs of the groups vary over a range of one to two standard deviations (15 to 30 IQ points), depending on the method of status classification (Eells, et al., 1951). This relationship between SES and IQ constitutes one of the most substantial and least disputed facts in psychology and education.

The fact that intelligence is correlated with occupational status can hardly be surprising in any society that supports universal public education. The educational system and occupational hierarchy act as an intellectual "screening" process, far from perfect, to be sure, but discriminating enough to create correlations of the magnitude just reported. If each generation is roughly sorted out by these "screening" processes along an intelligence continuum, and if, as has already been pointed out, the phenotype-genotype correlation for IQ is of the order of .80 to .90, it is almost inevitable that this sorting process will make for genotypic as well as phenotypic differences among social classes. It is therefore most unlikely that groups differing in SES would not also differ, on the average, in their genetic endowment of intelligence. In reviewing the relevant evidence, the British geneticist, C. O. Carter (1966, p. 192) remarked, "Sociologists who doubt this show more ingenuity than judgment." Sociologist Bruce Eckland (1967) has elaborately spelled out the importance of genetic factors for understanding social class differences.

Few if any students of this field today would regard socioeconomic status per se as an environmental variable that primarily *causes* IQ differences. Intellectual differences between SES groups have hereditary, environmental, and interaction components. Environmental factors associated with SES differences apparently are not a major *independent* source of variance in intelligence. Identical twins separated in the first months of life and reared in widely differing social classes, for example, still show greater similarity in IQ than unrelated children reared together or than even siblings reared together (Burt, 1966). The IQs of children adopted in infancy show a much lower correlation with the SES of the adopting parents than do the IQs of children reared by their own parents (Leahy, 1935). The IQs of children who were reared in an orphanage from infancy and who had never known their biological parents show approximately the same correlation with their biological father's occupational status as found for children reared by

their biological parents (.23 *vs* .24) (Lawrence, 1931). The correlation between the IQs of children adopted in infancy and the educational level of their biological mothers is close to that of children reared by their own mothers (.44), while the correlation between children's IQs and their adopting parents' educational level is close to zero (Honzik, 1957). Children of low and high SES show, on the average, an amount of regression from the parental IQ toward the mean of the general population that conforms to expectations from a simple polygenic model of the inheritance of intelligence (Burt, 1961). When siblings reared within the same family differ significantly in intelligence, those who are above the family average tend to move up the SES scale, and those who are below the family average tend to move down (Young & Gibson, 1965). It should also be noted that despite intensive efforts by psychologists, educators, and sociologists to devise tests intended to eliminate SES differences in measured intelligence, none of these efforts has succeeded (Jensen, 1968c). Theodosius Dobzhansky (1968a, p. 33), a geneticist, states that "There exist some occupations or functions for which only extreme genotypes are suitable." But surely this is not an all-or-nothing affair, and we would expect by the same reasoning that many different occupational skills, and not just those that are the most extreme, would favor some genotypes more than others. To be sure, genetic factors become more important at the extremes. Some minimal level of ability is required for learning most skills. But while you can teach almost anyone to play chess, or the piano, or to conduct an orchestra, or to write prose, you cannot teach everyone to be a Capablanca, a Paderewski, a Toscanini, or a Bernard Shaw. In a society that values and rewards individual talent and merit, genetic factors inevitably take on considerable importance.

SES differences, and race differences as well, are manifested not only as differences between group means, but also as differences in variance and in patterns of correlations among various mental abilities, even on tests which show no *mean* differences between SES groups (Jensen, 1968b).

Another line of evidence that SES IQ differences are not a superficial phenomenon is the fact of a negative correlation between SES and Developmental Quotient (DQ) (under two years of age) and an increasing positive correlation between SES and IQ (beyond two years of age), as shown in Figure 14 from a study by Nancy Bayley (1966). (All subjects in this study are Caucasian.) This relationship is especially interesting in view of the finding of a number of studies that there is a negative correlation between DQ and later IQ, an effect which is much more pronounced in boys than in girls and involves the motor more than the attentional-cognitive aspects of the DQ (Bayley, 1965b). Figure 14 shows that on

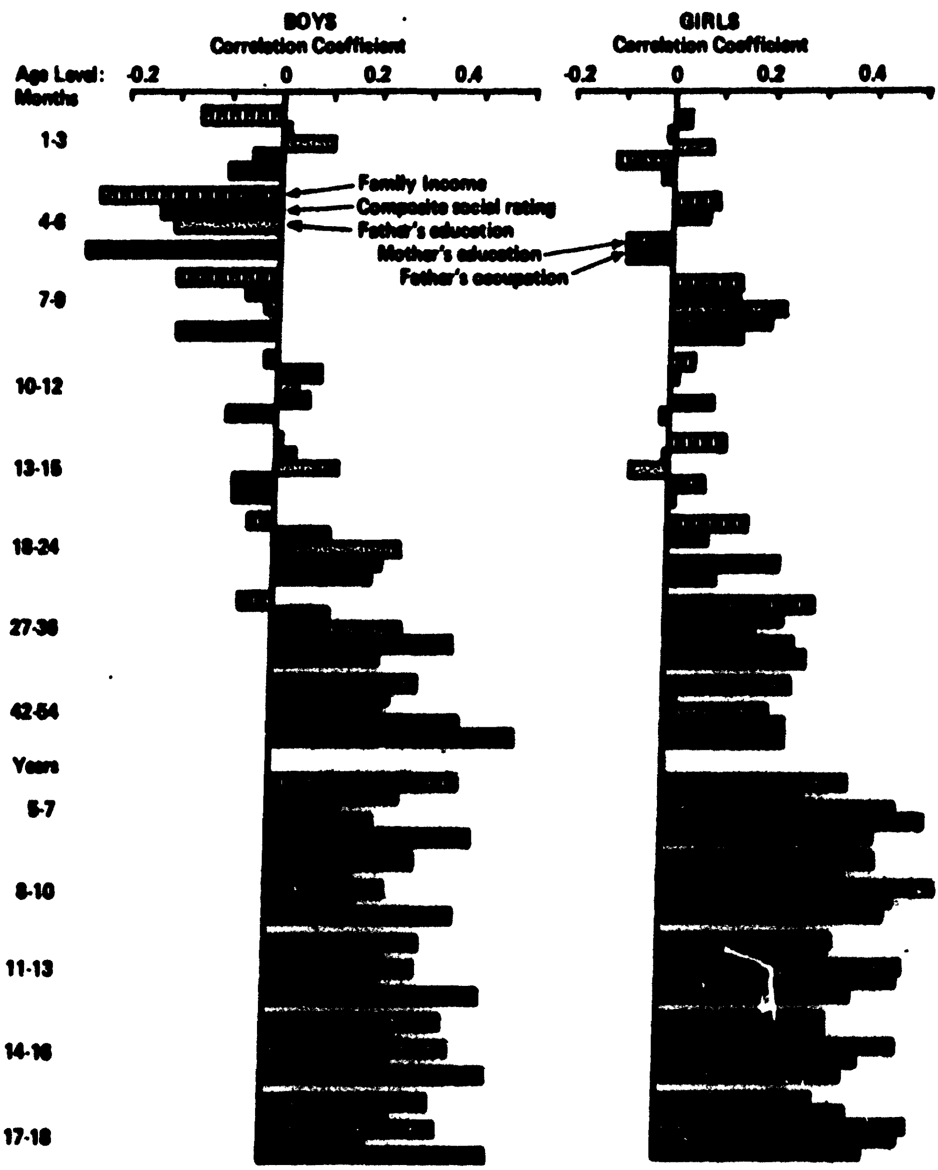


FIGURE 14.

Correlations between children's mental test scores, at 1 month to 18 years, and five indicators of parents' socioeconomic status at the time the children were born. (Bayley, 1966.)

infant developmental scales, lower SES children actually have a "head start" over higher SES children. But this trend is increasingly reversed at later ages as the tests become less motoric and are increasingly loaded with a cognitive or *g* factor.

Race Differences

The important distinction between the *individual* and the *population* must always be kept clearly in mind in any discussion of racial differences in mental abilities or any other behavioral characteristics. Whenever we select a person for some special educational purpose, whether for special instruction in a grade-school class for children with learning problems, or for a "gifted" class with an advanced curriculum, or for college attendance, or for admission to graduate training or a professional school, we are selecting an *individual*, and we are selecting him and dealing with him as an individual for reasons of his individuality. Similarly, when we employ someone, or promote someone in his occupation, or give some special award or honor to someone for his accomplishments, we are doing this to an individual. The variables of social class, race, and national origin are correlated so imperfectly with any of the valid criteria on which the above decisions should depend, or, for that matter, with any behavioral characteristic, that these background factors are irrelevant as a basis for dealing with individuals—as students, as employees, as neighbors. Furthermore, since, as far as we know, the full range of human talents is represented in all the major races of man and in all socioeconomic levels, it is unjust to allow the mere fact of an individual's racial or social background to affect the treatment accorded to him. All persons rightfully must be regarded on the basis of their individual qualities and merits, and all social, educational, and economic institutions must have built into them the mechanisms for insuring and maximizing the treatment of persons according to their individual behavior.

If a society completely believed and practiced the ideal of treating every person as an individual, it would be hard to see why there should be any problems about "race" per se. There might still be problems concerning poverty, unemployment, crime, and other social ills, and, given the will, they could be tackled just as any other problems that require rational methods for solution. But if this philosophy prevailed in practice, there would not need to be a "race problem."

The question of *race* differences in intelligence comes up not when we deal with individuals as individuals, but when certain identifiable *groups* or subcultures within the society are brought into comparison with one another *as groups or populations*. It is only when the groups are disproportionately represented in

what are commonly perceived as the most desirable and the least desirable social and occupational roles in a society that the question arises concerning average differences among groups. Since much of the current thinking behind civil rights, fair employment, and equality of educational opportunity appeals to the fact that there is a disproportionate representation of different racial groups in the various levels of the educational, occupational, and socioeconomic hierarchy, we are forced to examine all the possible reasons for this inequality among racial groups in the attainments and rewards generally valued by all groups within our society. To what extent can such inequalities be attributed to unfairness in society's multiple selection processes? ("Unfair" meaning that selection is influenced by intrinsically irrelevant criteria, such as skin color, racial or national origin, etc.) And to what extent are these inequalities attributable to really relevant selection criteria which apply equally to all individuals but at the same time select disproportionately between some racial groups because there exist, in fact, real average differences among the groups—differences in the population distributions of those characteristics which are indisputably relevant to educational and occupational performance? This is certainly one of the most important questions confronting our nation today. The answer, which can be found only through unfettered research, has enormous consequences for the welfare of all, particularly of minorities whose plight is now in the foreground of public attention. A preordained, doctrinaire stance with regard to this issue hinders the achievement of a scientific understanding of the problem. To rule out of court, so to speak, any reasonable hypotheses on purely ideological grounds is to argue that static ignorance is preferable to increasing our knowledge of reality. I strongly disagree with those who believe in searching for the truth by scientific means only under certain circumstances and eschew this course in favor of ignorance under other circumstances, or who believe that the results of inquiry on some subjects cannot be entrusted to the public but should be kept the guarded possession of a scientific elite. Such attitudes, in my opinion, represent a danger to free inquiry and, consequently, in the long run, work to the disadvantage of society's general welfare. "No holds barred" is the best formula for scientific inquiry. One does not decree beforehand which phenomena cannot be studied or which questions cannot be answered.

Genetic Aspects of Racial Differences. No one, to my knowledge, questions the role of environmental factors, including influences from past history, in determining at least some of the variance between racial groups in standard measures

of intelligence, school performance, and occupational status. The current literature on the culturally disadvantaged abounds with discussion—some of it factual, some of it fanciful—of how a host of environmental factors depresses cognitive development and performance. I recently co-edited a book which is largely concerned with the environmental aspects of disadvantaged minorities (Deutsch, Katz, & Jensen, 1968). But the possible importance of genetic factors in racial behavioral differences has been greatly ignored, almost to the point of being a tabooed subject, just as were the topics of venereal disease and birth control a generation or so ago.

My discussions with a number of geneticists concerning the question of a genetic basis of differences among races in mental abilities have revealed to me a number of rather consistently agreed-upon points which can be summarized in general terms as follows: Any groups which have been geographically or socially isolated from one another for many generations are practically certain to differ in their gene pools, and consequently are likely to show differences in any phenotypic characteristics having high heritability. This is practically axiomatic, according to the geneticists with whom I have spoken. Races are said to be "breeding populations," which is to say that matings within the group have a much higher probability than matings outside the group. Races are more technically viewed by geneticists as populations having different distributions of gene frequencies. These genetic differences are manifested in virtually every anatomical, physiological, and biochemical comparison one can make between representative samples of identifiable racial groups (Kuttner, 1967). There is no reason to suppose that the brain should be exempt from this generalization. (Racial differences in the relative frequencies of various blood constituents have probably been the most thoroughly studied so far.)

But what about behavior? If it can be measured and shown to have a genetic component, it would be regarded, from a genetic standpoint, as no different from other human characteristics. There seems to be little question that racial differences in genetically conditioned behavioral characteristics, such as mental abilities, should exist, just as physical differences. The real questions, geneticists tell me, are not whether there are or are not genetic racial differences that affect behavior, because there undoubtedly are. The proper questions to ask, from a scientific standpoint, are: What is the direction of the difference? What is the magnitude of the difference? And what is the significance of the difference—medically, socially, educationally, or from whatever standpoint that may be relevant to the characteristic in question? A difference is important only within a speci-

fic context. For example, one's blood type in the ABO system is unimportant until one needs a transfusion. And some genetic differences are apparently of no importance with respect to any context as far as anyone has been able to discover—for example, differences in the size and shape of ear lobes. The idea that all genetic differences have arisen or persisted only as a result of natural selection, by conferring some survival or adaptive benefit on their possessors, is no longer generally held. There appear to be many genetic differences, or polymorphisms, which confer no discernible advantages to survival.[6]

Negro Intelligence and Scholastic Performance. Negroes in the United States are disproportionately represented among groups identified as culturally or educationally disadvantaged. This, plus the fact that Negroes constitute by far the largest racial minority in the United States, has for many years focused attention on Negro intelligence. It is a subject with a now vast literature which has been quite recently reviewed by Dreger and Miller (1960, 1968) and by Shuey (1966), whose 578 page review is the most comprehensive, covering 382 studies. The basic data are well known: on the average, Negroes test about 1 standard deviation (15 IQ points) below the average of the white population in IQ, and this finding is fairly uniform across the 81 different tests of intellectual ability used in the studies reviewed by Shuey. This magnitude of difference gives a median overlap of 15 percent, meaning that 15 percent of the Negro population exceeds the white average. In terms of proportions of variance, if the numbers of Negroes and whites were equal, the differences *between* racial groups would account for 23 percent of the total variance, but—an important point—the differences *within* groups would account for 77 percent of the total variance. When gross socioeconomic level is controlled, the average difference reduces to about 11 IQ points (Shuey, 1966, p. 519), which, it should be recalled, is about the same spread as the average difference between siblings in the same family. So-called "culture-free" or "culture-fair" tests tend to give Negroes slightly lower scores, on the average, than more conventional IQ tests such as the Stanford-Binet and Wechsler scales. Also, as a group, Negroes perform somewhat more poorly on those subtests which tap abstract abilities. The majority of studies show that Negroes perform relatively better on verbal than on non-verbal intelligence tests.

In tests of scholastic achievement, also, judging from the massive data of the Coleman study (Coleman, et al., 1966), Negroes score about 1 standard devia-

[6] The most comprehensive and sophisticated discussion of the genic-behavior analysis of race differences that I have found is by Spuhler and Lindzey (1967).

tion (SD) below the average for whites and Orientals and considerably less than 1 SD below other disadvantaged minorities tested in the Coleman study—Puerto Rican, Mexican-American, and American Indian. The 1 SD decrement in Negro performance is fairly constant throughout the period from grades 1 through 12.

Another aspect of the distribution of IQs in the Negro population is their lesser variance in comparison to the white distribution. This shows up in most of the studies reviewed by Shuey. The best single estimate is probably the estimate based on a large normative study of Stanford-Binet IQs of Negro school children in five Southeastern states, by Kennedy, Van De Riet, and White (1963). They found the SD of Negro children's IQs to be 12.4, as compared with 16.4 in the white normative sample. The Negro distribution thus has only about 60 percent as much variance (i.e., SD^2) as the white distribution.

There is an increasing realization among students of the psychology of the disadvantaged that the discrepancy in their average performance cannot be completely or directly attributed to discrimination or inequalities in education. It seems not unreasonable, in view of the fact that intelligence variation has a large genetic component, to hypothesize that genetic factors may play a part in this picture. But such an hypothesis is anathema to many social scientists. The idea that the lower average intelligence and scholastic performance of Negroes could involve, not only environmental, but also genetic, factors has indeed been strongly denounced (e.g., Pettigrew, 1964). But it has been neither contradicted nor discredited by evidence.

The fact that a reasonable hypothesis has not been rigorously proved does not mean that it should be summarily dismissed. It only means that we need more appropriate research for putting it to the test. I believe such definitive research is entirely possible but has not yet been done. So all we are left with are various lines of evidence, no one of which is definitive alone, but which, viewed all together, make it a not unreasonable hypothesis that genetic factors are strongly implicated in the average Negro-white intelligence difference. The preponderance of the evidence is, in my opinion, less consistent with a strictly environmental hypothesis than with a genetic hypothesis, which, of course, does not exclude the influence of environment or its interaction with genetic factors.

We can be accused of superficiality in our thinking about this issue, I believe, if we simply dismiss a genetic hypothesis without having seriously thought about the relevance of typical findings such as the following:

Failure to Equate Negroes and Whites in IQ and Scholastic Ability. No one has yet produced any evidence based on a properly controlled study to show that rep-

resentative samples of Negro and white children can be equalized in intellectual ability through statistical control of environment and education.

Socioeconomic Level and Incidence of Mental Retardation. Since in no category of socioeconomic status (SES) are a majority of children found to be retarded in the technical sense of having an IQ below 75, it would be hard to claim that the degree of environmental deprivation typically associated with lower-class status could be responsible for this degree of mental retardation. An IQ less than 75 reflects more than a lack of cultural amenities. Heber (1968) has estimated on the basis of existing evidence that IQs below 75 have a much higher incidence among Negro than among white children at every level of socioeconomic status, as shown in Table 3. In the two highest SES categories the estimated proportions of Negro and white children with IQs below 75 are in the ratio of 13.6 to 1. If

TABLE 3

Estimated Prevalence of Children With IQs Below 75, by Socioeconomic Status (SES) and Race Given as Percentages (Heber, 1968)

SES	*White*	*Negro*
High 1	0.5	3.1
2	0.8	14.5
3	2.1	22.8
4	3.1	37.8
Low 5	7.8	42.9

environmental factors were mainly responsible for producing such differences, one should expect a lesser Negro-white discrepancy at the upper SES levels. Other lines of evidence also show this not to be the case. A genetic hypothesis, on the other hand, would predict this effect, since the higher SES Negro offspring would be regressing to a lower population mean than their white counterparts in SES, and consequently a larger proportion of the lower tail of the distribution of genotypes for Negroes would fall below the value that generally results in phenotypic IQs below 75.

A finding reported by Wilson (1967) is also in line with this prediction. He obtained the mean IQs of a large representative sample of Negro and white children in a California school district and compared the two groups within each of four social class categories: (1) professional and managerial, (2) white collar, (3)

skilled and semiskilled manual, and (4) lower class (unskilled, unemployed, or welfare recipients). The mean IQ of Negro children in the first category was 15.5 points below that of the corresponding white children in SES category 1. But the Negro mean for SES 1 was also 3.9 points below the mean of white children in SES category 4. (The IQs of white children in SES 4 presumably have "regressed" upward toward the mean of the white population.)

Wilson's data are not atypical, for they agree with Shuey's (1966, p. 520) summarization of the total literature up to 1965 on this point. She reports that in all the studies which grouped subjects by SES, upper-status Negro children average 2.6 IQ points *below* the low-status whites. Shuey comments: "It seems improbable that upper and middle-class colored children would have no more culture opportunities provided them than white children of the lower and lowest class."

Duncan (1968, p. 69) also has presented striking evidence for a much greater "regression-to-the-mean" (from parents to their children) for high status occupations in the case of Negroes than in the case of whites. None of these findings is at all surprising from the standpoint of a genetic hypothesis, of which an intrinsic feature is Galton's "law of filial regression." While the data are not necessarily inconsistent with a possible environmental interpretation, they do seem more puzzling in terms of strictly environmental causation. Such explanations often seem intemperately strained.

Inadequacies of Purely Environmental Explanations. Strictly environmental explanations of group differences tend to have an ad hoc quality. They are usually plausible for the situation they are devised to explain, but often they have little generality across situations, and new ad hoc hypotheses have to be continually devised. Pointing to environmental differences between groups is never sufficient in itself to infer a causal relationship to group differences in intelligence. To take just one example of this tendency of social scientists to attribute lower intelligence and scholastic ability to almost any environmental difference that seems handy, we can look at the evidence regarding the effects of "father absence." Since the father is absent in a significantly larger proportion of Negro than of white families, the factor of "father absence" has been frequently pointed to in the literature on the disadvantaged as one of the causes of Negroes' lower performance on IQ tests and in scholastic achievement. Yet the two largest studies directed at obtaining evidence on this very point—the only studies I have seen that are methodologically adequate—both conclude that the factor of "father absence"

versus "father presence" makes no independent contribution to variance in intelligence or scholastic achievement. The sample sizes were so large in both of these studies that even a very slight degree of correlation between father-absence and the measures of cognitive performance would have shown up as statistically significant. Coleman (1966, p. 506) concluded: "Absence of a father in the home did not have the anticipated effect on ability scores. Overall, pupils without fathers performed at approximately the same level as those with fathers—although there was some variation between groups" (groups referring to geographical regions of the U.S.). And Wilson (1957, p. 177) concluded from his survey of a California school district: "Neither our own data nor the preponderance of evidence from other research studies indicate that father presence or absence, *per se*, is related to school achievement. While broken homes reflect the existence of social and personal problems, and have some consequence for the development of personality, broken homes do not have any systematic effect on the overall level of school success."

The nationwide Coleman study (1966) included assessments of a dozen environmental variables and socioeconomic indices which are generally thought to be major sources of environmental influence in determining individual and group differences in scholastic performance—such factors as: reading material in the home, cultural amenities in the home, structural integrity of the home, foreign language in the home, preschool attendance, parents' education, parents' educational desires for child, parents' interest in child's school work, time spent on homework, child's self-concept (self-esteem), and so on. These factors are all correlated—in the expected direction—with scholastic performance within each of the racial or ethnic groups studied by Coleman. Yet, interestingly enough, they are not systematically correlated with differences *between* groups. For example, by far the most environmentally disadvantaged groups in the Coleman study are the American Indians. On every environmental index they average *lower* than the Negro samples, and overall their environmental rating is about as far below the Negro average as the Negro rating is below the white average. (As pointed out by Kuttner [1968, p. 707], American Indians are much more disadvantaged than Negroes, or any other minority groups in the United States, on a host of other factors not assessed by Coleman, such as income, unemployment, standards of health care, life expectancy, and infant mortality.) Yet the American Indian ability and achievement test scores average about half a standard deviation higher than the scores of Negroes. The differences were in favor of the Indian children on each of the four tests used by Coleman: non-verbal intelligence, ver-

bal intelligence, reading comprehension, and math achievement. If the environmental factors assessed by Coleman are the major determinants of Negro-white differences that many social scientists have claimed they are, it is hard to see why such factors should act in reverse fashion in determining differences between Negroes and Indians, especially in view of the fact that *within* each group the factors are significantly correlated in the expected direction with achievement.

Early Developmental Differences. A number of students of child development have noted the developmental precocity of Negro infants, particularly in motoric behavior. Geber (1958) and Geber and Dean (1957) have reported this precocity also in African infants. It hardly appears to be environmental, since it is evident in nine-hour-old infants. Cravioto (1966, p. 78) has noted that the Gesell tests of infant behavioral development, which are usually considered suitable only for children over four weeks of age, "can be used with younger African, Mexican, and Guatemalan infants, since their development at two or three weeks is similar to that of Western European infants two or three times as old." Bayley's (1965a) study of a representative sample of 600 American Negro infants up to 15 months of age, using the Bayley Infant Scales of Mental and Motor Development, also found Negro infants to have significantly higher scores than white infants in their first year. The difference is largely attributable to the motor items in the Bayley test. For example, about 30 percent of white infants as compared with about 60 percent of Negro infants between 9 and 12 months were able to "pass" such tests as "pat-a-cake" muscular coordination, and ability to walk with help, to stand alone, and to walk alone. The highest scores for any group on the Bayley scales that I have found in my search of the literature were obtained by Negro infants in the poorest sections of Durham, North Carolina. The older siblings of these infants have an average IQ of about 80. The infants up to 6 months of age, however, have a Developmental Motor Quotient (DMQ) nearly one standard deviation above white norms and a Developmental IQ (i.e., the non-motor items of the Bayley scale) of about half a standard deviation above white norms (Durham Education Improvement Program, 1966-67, a, b).

The DMQ, as pointed out previously, correlates negatively in the white population with socioeconomic status and with later IQ. Since lower SES Negro and white school children are more alike in IQ than are upper SES children of the two groups (Wilson, 1967), one might expect greater DMQ differences in favor of Negro infants in high socioeconomic Negro and white samples than in low socioeconomic samples. This is just what Walters (1967) found. High SES Negro in-

77-948 O - 72 - 11

fants significantly exceeded whites in total score on the Gesell developmental schedules at 12 weeks of age, while low SES Negro and white infants did not differ significantly overall. (The only difference, on a single subscale, favored the white infants.)

It should also be noted that developmental quotients are usually depressed by adverse prenatal, perinatal, and postnatal complications such as lack of oxygen, prematurity, and nutritional deficiency.

Another relationship of interest is the finding that the negative correlation between DMQ and later IQ is higher in boys than in girls (Bayley, 1966, p. 127) Bronfenbrenner (1967, p. 912) cites evidence which shows that Negro boys perform relatively less well in school than Negro girls; the sex difference is much greater than is found in the white population. Bronfenbrenner (1967, p. 913) says, "It is noteworthy that these sex differences in achievement are observed among Southern as well as Northern Negroes, are present at every socioeconomic level, and tend to increase with age."

Physiological Indices. The behavioral precocity of Negro infants is also paralleled by certain physiological indices of development. For example, x-rays show that bone development, as indicated by the rate of ossification of cartilege, is more advanced in Negro as compared with white babies of about the same socioeconomic background, and Negro babies mature at a lower birth-weight than white babies (Naylor & Myrianthopoulos, 1967, p. 81).

It has also been noted that brain wave patterns in African newborn infants show greater maturity than is usually found in the European newborn child (Nilson & Dean, 1959). This finding especially merits further study, since there is evidence that brain waves have some relationship to IQ (Medical World News, 1968), and since at least one aspect of brain waves—the visually evoked potential—has a very significant genetic component, showing a heritability of about .80 (uncorrected for attenuation) (Dustman & Beck, 1965).

Magnitude of Adult Negro-White Differences. The largest sampling of Negro and white intelligence test scores resulted from the administration of the Armed Forces Qualification Test (AFQT) to a national sample of over 10 million men between the ages of 18 and 26. As of 1966, the overall failure rate for Negroes was 68 percent as compared with 19 percent for whites (*U.S. News and World Report*, 1966). (The failure cut-off score that yields these percentages is roughly equivalent to a Stanford-Binet IQ of 86.) Moynihan (1965) has estimated that during the same period in which the AFQT was administered to these large representa-

tive samples of Negro and white male youths, approximately one-half of Negro families could be considered as middle-class or above by the usual socioeconomic criteria. So even if we assumed that all of the lower 50 percent of Negroes on the SES scale failed the AFQT, it would still mean that at least 36 percent of the middle SES Negroes failed the test, a failure rate almost twice as high as that of the white population for all levels of SES.

Do such findings raise any question as to the plausibility of theories that postulate exclusively environmental factors as sufficient causes for the observed differences?

Why Raise Intelligence?

If the intelligence of the whole population increased and our IQ tests were standardized anew, the mean IQ would again be made equal to 100, which, by definition, is the average for the population. Thus, in order to speak sensibly of raising intelligence we need an absolute frame of reference, and for simplicity's sake we will use the *present* distribution of IQ as our reference scale. Then it will not be meaningless to speak of the average IQ of the population shifting to values other than 100.

Would there be any real advantage to shifting the entire distribution of intelligence upward? One way to answer this question is to compare the educational attainments of children in different schools whose IQ distributions center around means of, say, 85, 100, and 115. As pointed out earlier, there is a relationship between educational attainments and the occupations that are open to individuals on leaving school. Perusal of the want-ads in any metropolitan newspaper reveals that there are extremely few jobs advertised which are suitable to the level of education and skills typically found below IQs of 85 or 90, while we see day after day in the want-ads hundreds of jobs which call for a level of education and skills typically found among school graduates with IQs above 110. These jobs go begging to be filled. The fact is, there are not nearly enough minimally qualified persons to fill them.

One may sensibly ask the question whether our collective national intelligence is adequate to meet the growing needs of our increasingly complex industrial society. In a bygone era, when the entire population's work consisted almost completely of gathering or producing food by primitive means, there was little need for a large number of persons with IQs much above 100. Few of the jobs that had to be done at that time required the kinds of abstract intelligence and

academic training which are now in such seemingly short supply in relation to the demand in our modern society. For many years the criterion for mental retardation was an IQ below 70. In recent years the National Association for Mental Retardation has raised the criterion to an IQ of 85, since an increasing proportion of persons of more than 1 standard deviation below the average in IQ are unable to get along occupationally in today's world. Persons with IQs of 85 or less are finding it increasingly difficult to get jobs, any jobs, because they are unprepared, for whatever reason, to do the jobs that need doing in this industrialized, technological economy. Unless drastic changes occur—in the population, in educational outcomes, or in the whole system of occupational training and selection—it is hard to see how we can avoid an increase in the rate of the so-called "hard-core" unemployed. It takes more knowledge and cleverness to operate, maintain, or repair a tractor than to till a field by hand, and it takes more skill to write computer programs than to operate an adding machine. And apparently the trend will continue.

It has been argued by Harry and Margaret Harlow that "human beings in our world today have no more, or little more, than the absolute minimal intellectual endowment necessary for achieving the civilization we know today" (Harlow & Harlow, 1962, p. 34). They depict where we would probably be if man's average genetic endowment for intelligence had never risen above the level corresponding to IQ 75: "... the geniuses would barely exceed our normal or average level; comparatively few would be equivalent in ability to our average high school graduates. There would be no individuals with the normal intellectual capacities essential for making major discoveries, and there could be no civilization as we know it."

It may well be true that the kind of ability we now call intelligence was needed in a certain percentage of the human population for our civilization to have arisen. But while a small minority—perhaps only one or two percent—of highly gifted individuals were needed to advance civilization, the vast majority were able to assimilate the consequences of these advances. It may take a Leibnitz or a Newton to invent the calculus, but almost any college student can learn it and use it.

Since intelligence (meaning *g*) is not the whole of human abilities, there may be some fallacy and some danger in making it the *sine qua non* of fitness to play a productive role in modern society. We should not assume certain ability requirements for a job without establishing these requirements as a fact. How often do employment tests, Civil Service examinations, the requirement of a high school

diploma, and the like, constitute hurdles that are irrelevant to actual performance on the job for which they are intended as a screening device? Before going overboard in deploring the fact that disadvantaged minority groups fail to clear many of the hurdles that are set up for certain jobs, we should determine whether the educational and mental test barriers that stand at the entrance to many of these employment opportunities are actually relevant. They may be relevant only in the correlational sense that the test predicts success on the job, in which case we should also know whether the test measures the ability actually required on the job or measures only characteristics that happen to be correlated with some third factor which is really essential for job performance. Changing people in terms of the really essential requirements of a given job may be much more feasible than trying to increase their abstract intelligence or level of performance in academic subjects so that they can pass irrelevant tests.

IQ Gains from Environmental Improvement

As was pointed out earlier, since the environment acts as a threshold variable with respect to IQ, an overall increase in IQ in a population in which a great majority are above the threshold, such that most of the IQ variance is due to heredity, could not be expected to be very large if it had to depend solely upon improving the environment of the economically disadvantaged. This is not to say that such improvement is not to be desired for its own sake or that it would not boost the educational potential of many disadvantaged children. An unrealistically high upper limit of what one could expect can be estimated from figures given by Schwebel (1968, p. 210). He estimates that 26 percent of the children in the population can be called environmentally deprived. He estimates the frequencies of their IQs in each portion of the IQ scale; their distribution is skewed, with higher frequencies in the lower IQ categories and an overall mean IQ of 90. Next, he assumes we could add 20 points to each deprived child's IQ by giving him an abundant environment. (The figure of 20 IQ points comes from Bloom's [1964, p. 89] estimate that the effect of extreme environments on intelligence is about 20 IQ points.) The net effect of this 20-point boost in the IQ of every deprived child would be an increase in the population's IQ from 100 to 105. But this seems to be an unrealistic fantasy. For if it were true that the IQs of the deprived group could be raised 20 points by a good environment, and if Schwebel's estimate of 26 percent correctly represents the incidence of deprivation, then the deprived children would be boosted to an average IQ of 110, which is 7 points higher than the mean of 103 for the non-deprived population! There

is no reason to believe that the IQs of deprived children, given an environment of abundance, would rise to a higher level than the already privileged children's IQs. The overall boost in the population IQ would probably be more like 1 or 2 IQ points rather than 5. (Another anomaly of Schwebel's "analysis" is that after a 20-point IQ boost is granted to the deprived segment of the population, the only persons left in the mentally retarded range are the non-deprived, with 7 percent of them below IQ 80 as compared with zero percent of the deprived!)

Fewer persons, however, are seriously concerned about whether or not we could appreciably boost the IQ of the population as a whole. A more feasible and urgent goal is to foster the educational and occupational potential of the disadvantaged segment of the population. The pursuit of this aim, of course, must involve advances not only in education, but in public health, in social services, and in welfare and employment practices. In considering all feasible measures, one must also take inventory of forces that may be working against the accomplishment of amelioration. We should not overlook the fact that social and economic conditions not only have direct environmental effects, but indirectly can have biological consequences as well, consequences that could oppose attempts to improve the chances of the disadvantaged to assume productive roles in society.

Possible Dysgenic Trends

In one large midwestern city it was found that one-third of all the children in classes for the mentally retarded (IQ less than 75) came from one small area of the city comprising only five percent of the city's population (Heber, 1968). A representative sample of 88 mothers having at least one school-age child in this neighborhood showed an average of 7.6 children per mother. In families of 8 or more, nearly half the children over 12 years of age had IQs below 75 (Heber, Dener, & Conry, 1968). The authors note that not all low SES families contributed equally to the rate of mental retardation in this area; certain specifiable families had a greatly disproportionate number of retarded children. Mothers with IQs below 80, for example, accounted for over 80 percent of the children with IQs under 80. Completely aside from the hereditary implications, what does this mean in view of studies of foster children which show that the single most important factor in the child's *environment* with respect to his intellectual development is his foster mother's IQ? This variable has been shown to make the largest *independent* contribution to variance in children's IQs of any environmental factor (Burks, 1928). If the children in the neighborhoods studied by

Heber, which are typical of the situation in many of our large cities, have the great disadvantage of deprived environments, is it inappropriate to ask the same question that Florence Goodenough (1940, p. 329) posed regarding causal factors in retarded Tennessee mountain children: "*Why* are they so deprived?" When a substantial proportion of the children in a community suffer a deplorable environment, one of the questions we need to answer is who creates their environment? Does not the genetic × environment interaction work both ways, the genotype to some extent making its own environment and that of its progeny?

In reviewing evidence from foster home studies on environmental amelioration of IQs below 75 (the range often designated as indicating cultural-familial retardation) Heber, Dever, and Conry (1968, p. 17) state: "The conclusion that changes in the living environment can cause very large increments in IQ *for the cultural-familial retardate* is not warranted by these data."

What is probably the largest study every made of familial influences in mental retardation (defined in this study as IQ less than 70) involved investigation of more than 80,000 relatives of a group of mentally retarded persons by the Dight Institute of Genetics, University of Minnesota (Reed & Reed, 1965). From this large-scale study, Sheldon and Elizabeth Reed estimated that about 80 percent of mentally retarded (IQ less than 70) persons in the United States have a retarded parent or a normal parent who has a retarded sibling. The Reeds state: "One inescapable conclusion is that the transmission of mental retardation from parent to child is by far the most important *single* factor in the persistence of this social misfortune" (p. 48). "The transmission of mental retardation from one generation to the next, should, therefore, receive much more critical attention than it has in the past. It seems fair to state that this problem has been largely ignored on the assumption that if our social agencies function better, that if everyone's environment were improved sufficiently, then mental retardation would cease to be a major problem" (p. 77).

An interesting sidelight of the Reeds' study is the finding that in a number of families in which one or both parents had IQs below 70 and in which the environment they provided their children was deplorably deprived, there were a few children of average and superior IQ (as high as 130 or above) and superior scholastic performance. From a genetic standpoint the occurrence of such children would be expected. It is surprising from a strictly environmental standpoint. But, even though some proportion of the children of retarded parents are obviously intellectually well endowed, who would wish upon them the kind of environment typically provided by retarded parents? An investigation conducted

in Denmark concluded that "...it is a very severe psychical trauma for a normally gifted child to grow up in a home where the mother is mentally deficient" (Jepsen & Bredmose, 1956, p. 209). Have we thought sufficiently of the rights of children—of their right to be born with fair odds against being mentally retarded, not to have a retarded parent, and with fair odds in favor of having the genetic endowment needed to compete on equal terms with the majority of persons in society? Can we reasonably and humanely oppose such rights of millions of children as yet not born?

Is Our National IQ Declining? It has long been known that there is a substantial negative correlation (averaging about -.30 in various studies) between intelligence and family size and between social class and family size (Anastasi, 1956). Children with many siblings, on the average, have lower IQs than children in small families, and the trend is especially marked for families of more than five (Gottesman, 1968). This fact once caused concern in the United States, and even more so in Britain, because of its apparent implication of a declining IQ in the population. If more children are born to persons in the lower half of the intelligence distribution, one would correctly predict a decline in the average IQ of the population. In a number of large-scale studies addressed to the issue in Britain and the United States some 20 years ago, no evidence was found for a general decline in IQ (Duncan, 1952). The paradox of the apparent failure of the genetic prediction to be manifested was resolved to the satisfaction of most geneticists by three now famous studies, one by Higgins, Reed, and Reed (1962), the others by Bajema (1963, 1966). All previous analyses had been based on IQ comparisons of children having different numbers of siblings, and this was their weakness. The data needed to answer the question properly consist of the average number of children born to *all* individuals at every level of IQ. It was found in the three studies that if persons with very low IQs married and had children, they typically had a large number of children. *But*—it was also found that relatively few persons in the lower tail of the IQ distribution ever married or produced children, and so their reproduction rate is more than counterbalanced by persons at the upper end of the IQ scale, nearly all of whom marry and have children. The data of these studies are shown in Figure 15.

In my opinion these studies are far from adequate to settle this issue and thus do not justify complacency. They cannot be generalized much beyond the particular generation which the data represent or to other than the white population on which these studies were based. The population sampled by Bajema

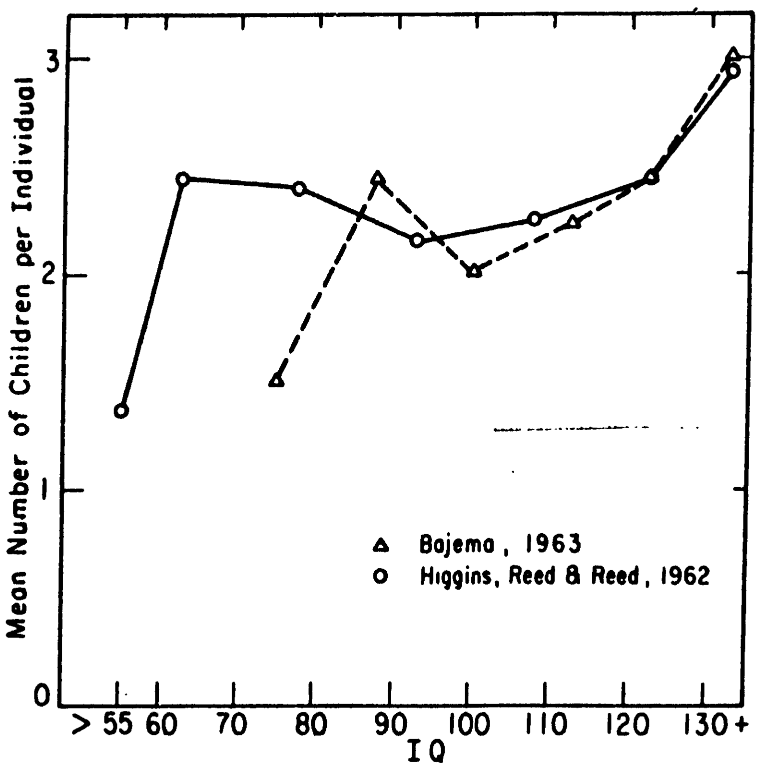


FIGURE 15.

Mean number of children per adult individual (including those who are childless) at each level of IQ, in two samples of white American populations. Note in each sample the bimodal relationship between fertility and IQ.

(1963, 1966), for example, consisted of native-born American whites, predominantly Protestant, with above-average educational attainments, living all or most of their lives in an urban environment, and having most of their children before World War II. Results from a study of this population cannot be confidently generalized to other, quite dissimilar segments of our national population. The relationship between reproductive rate and IQ found by Bajema and by Higgins et al. may very well not prevail in every population group. Thus the evidence to date has not nullified the question of whether dysgenic trends are operating in some sectors.

If this conclusion is not unwarranted, then our lack of highly relevant information on this issue with respect to our Negro population is deplorable, and no one should be more concerned about it than the Negro community itself. Certain census statistics suggest that there might be forces at work which could create and widen the genetic aspect of the average difference in ability between the Negro and white populations in the United States, with the possible consequence that the improvement of educational facilities and increasing equality of opportunity will have a *decreasing* probability of producing equal achievement or continuing gains in the Negro population's ability to compete on equal terms. The relevant statistics have been presented by Moynihan (1966). The differential birthrate, as a function of socioeconomic status, is greater in the Negro than in the white population. The data showing this relationship for one representative age group from the U.S. Census of 1960 are presented in Figure 16.

Negro middle- and upper-class families have fewer children than their white counterparts, while Negro lower-class families have more. In 1960, Negro women of ages 35 to 44 married to unskilled laborers had 4.7 children as compared with 3.8 for non-Negro women in the same situation. Negro women married to professional or technical workers had only 1.9 children as compared with 2.4 for white women in the same circumstances. Negro women with annual incomes below $2000 averaged 5.3 children. The poverty rate for families with 5 or 6 children is 3½ times as high as that for families with one or two children (Hill & Jaffe, 1966). That these figures have some relationship to intellectual ability is seen in the fact that 3 out of 4 Negroes failing the Armed Forces Qualification Test come from families of four or more children.

Another factor to be considered is average generation time, defined as the number of years it takes for the parent generation to reproduce its own number. This period is significantly less in the Negro than in the white population. Also, as noted in the study of Bajema (1966), generation length is inversely related to educational attainment and occupational status; therefore a group with shorter generation length is more likely subject to a possible dysgenic effect.

Much more thought and research should be given to the educational and social implications of these trends for the future. Is there a danger that current welfare policies, unaided by eugenic foresight, could lead to the genetic enslavement of a substantial segment of our population? The possible consequences of our failure seriously to study these questions may well be viewed by future generations as our society's greatest injustice to Negro Americans.

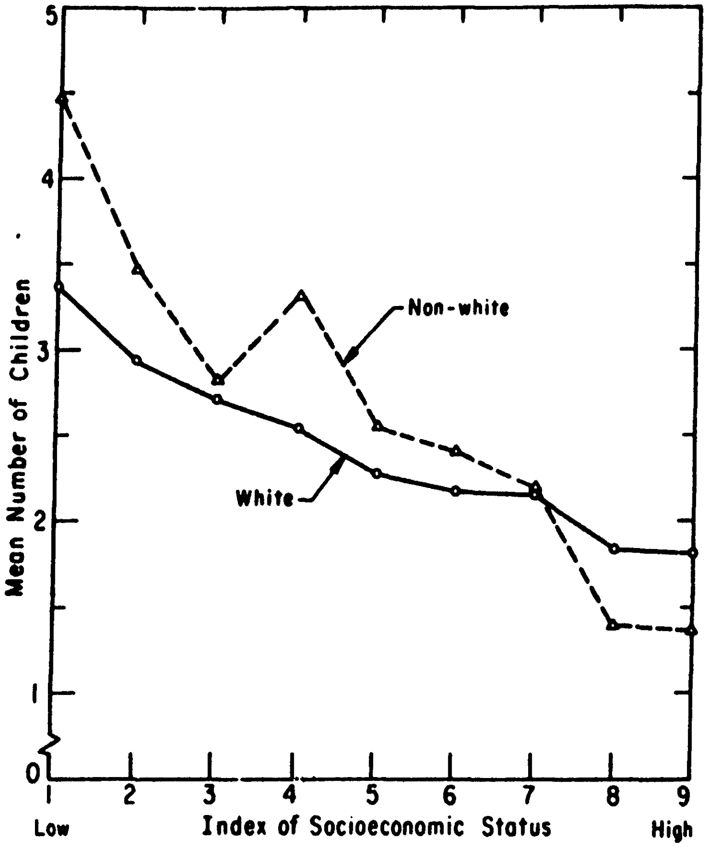


FIGURE 16.

Average number of children per woman 25 to 29 years of age, married once, with husband present, by race and socioeconomic status. From 1960 U.S. Census. (After Mitra, 1966.)

Intensive Educational Intervention

We began with mention of several of the major compensatory education programs and their general lack of success in boosting the scholastic performance of disadvantaged children. It has been claimed that such mammoth programs have not been adequately pinpointed to meeting specific, fine-grained cultural and cognitive needs of these children and therefore should not be expected to produce the gains that could result from more intensive and more carefully fo-

cused programs in which maximum cultural enrichment and instructional ingenuity are lavished on a small group of children by a team of experts.

The scanty evidence available seems to bear this out. While massive compensatory programs have produced no appreciable gains in intelligence or achievement (as noted on pp. 2-3), the majority of small-scale experiments in boosting the IQ and educational performance of disadvantaged children have produced significant gains. It is interesting that the magnitude of claimed gains generally decreases as one proceeds from reports in the popular press, to informal verbal reports heard on visits at research sites and in private correspondence, to papers read at meetings, to published papers without presentation of supporting data, and to published papers with supporting data. I will confine my review to some of the major studies in the last category.

First, some general observations.

Magnitude of Gains. The magnitude of IQ and scholastic achievement gains resulting from enrichment and cognitive stimulation programs authentically range between about 5 and 20 points for IQs, and between about one-half to two standard deviations for specific achievement measures (reading, arithmetic, spelling, etc.). Heber (1968) reviewed 29 intensive preschool programs for disadvantaged children and found they resulted in an average gain in IQ (at the time of children's leaving the preschool program) of between 5 and 10 points; the average gain was about the same for children whose initial IQs were below 90 as for those of 90 and above.

The amount of gain is related to several factors. The intensity and specificity of the instructional aspects of the program seem to make a difference. Ordinary nursery school attendance, with a rather diffuse enrichment program but with little effort directed at development of specific cognitive skills, generally results in a gain of 5 or 6 IQ points in typical disadvantaged preschoolers. If special cognitive training, especially in verbal skills, is added to the program, the average gain is about 10 points—slightly more or less depending on the amount of verbal content in the tests. Average gains rarely go above this, but when the program is extended beyond the classroom into the child's home, and there is intensive instruction in specific skills under short but highly attention-demanding daily sessions, as in the Bereiter-Engelmann program (1966), about a third of the children have shown gains of as much as 20 points.

Average gains of more than 10 or 15 points have not been obtained on any sizeable groups or been shown to persist or to be replicable in similar groups,

although there have been claims that average gains of 20 or more points can be achieved by removing certain cultural and attitudinal barriers to learning. The actual evidence, however, warrants the caution expressed by Bereiter and Engelmann (1966, p. 7): " 'Miracle cures' of this kind are sometimes claimed to work with disadvantaged children, as when a child is found to gain 20 points or so in IQ after a few months of preschool experience. Such enormous gains, however, are highly suspect to anyone who is familiar with mental measurements. It is a fair guess that the child could have done as well on the first test except that he misinterpreted the situation, was frightened or agitated, or was not used to responding to instructions. Where genuine learning is concerned, enormous leaps simply do not occur, and leaps of any kind do not occur without sufficient cause."

The initial IQ on entering also has some effect, and this fact may be obscured if various studies are coarsely grouped. Bereiter and Engelmann (1966, p. 16), in analyzing results from eight different preschools for culturally disadvantaged children that followed traditional nursery school methods, concluded that the children's average gain in IQ is *half* the way from their initial IQ level to the normal level of 100. This rule was never more than 2 points in error for the studies reviewed. This same amount of IQ gain is generally noted in disadvantaged children during their first year in regular kindergarten (Brison, 1967, p. 8).

I have found no evidence of comparable gains in non-disadvantaged children. Probably the exceedingly meager gains in some apparently excellent preschool programs for the "disadvantaged" are attributable to the fact that the children in them did not come from a sufficiently deprived home background. Such can be the case when the children are admitted to the program on the basis of "self-selection" by their parents. Parents who seek out a nursery school or volunteer their children for an experimental preschool are more apt to have provided their children with a somewhat better environment than would be typical for a randomly selected group of disadvantaged children. This seems to have been the case in Martin Deutsch's intensive preschool enrichment program at the Institute of Developmental Studies in New York (Powledge, 1967). Both the experimental group (E) and the self-selected control groups (C_{ss}) were made up of Negro children from a poor neighborhood in New York City whose parents applied for their admission to the program. The E group received intensive educational attention in what is overall the most comprehensive and elaborate enrichment program I know of. The C_{ss} group, of course, received no enriched education.

The initial average Stanford-Binet IQs of the E and C_u groups were 93.32 and 94.69, respectively. After two years in the enrichment program, the E group had a mean IQ of 95.53 and the C_u group had 96.52. Both pre- and post-test differences are nonsignificant. The enrichment program continued for a third year through the first grade. For the children in the E group who had had three years of enrichment, there was a significant gain over the C group of 8 months in reading achievement by the end of first grade, a score above national norms. This result is in keeping with the general finding that enrichment shows a greater effect on scholastic achievement than on IQ per se.

Many studies have employed no control group selected on exactly the same basis as the experimental group. This makes it virtually impossible to evaluate the effect of the treatment on pre-test—post-test gain, and the problem is made more acute by the fact that enrichment studies often pick their subjects on the basis of their being below the average IQ of the population of disadvantaged children from which they are selected. This makes statistical regression a certainty—the group's mean will increase by an appreciable amount because of the imperfect correlation between test-retest scores over, say, a one-year interval. Since this correlation is known to be considerably lower in younger than in older children, there will be considerably greater "gain" due to regression for younger groups of children. The net results of selecting especially backward children on the basis of IQ is that a gain in IQ can be predicted which is not at all attributable to the educational treatment given to the children. Studies using control groups nearly always show this gain in the control group, and only by subtracting the control group's gain from the experimental group's gain can we evaluate the magnitude of the treatment effect. Only the gain over and above that attributable to regression really counts.

Still another factor is involved in the inverse relationship generally found between children's age and the size of IQ gains in an enrichment program. Each single item gotten right in a test like the Stanford-Binet adds increasingly smaller increments to the IQ as children get older. Each Stanford-Binet test item, for example, is worth two months of mental age. At four years of age getting just two additional items right will boost an IQ of 85 up to 93. The same absolute amount of improvement in test performance at 10 years of age would boost an IQ of 85 up to only 88. The typical range of gains found in preschool enrichment programs, in the age range of 4 to 6, are about what would be expected from passing an additional two to four items in the Stanford-Binet. This amount of gain should not be surprising on a test which, for this age range, consists of items

rather similar to the materials and activities traditionally found in nursery schools—blocks, animal pictures, puzzles, bead stringing, copying drawings, and the like. I once visited an experimental preschool using the Stanford-Binet to assess pre-test—post-test gains, in which some of the Stanford-Binet test materials were openly accessible to the children throughout their time in the school as part of the enrichment paraphernalia. Years ago Reymert and Hinton (1940) noted this "easy gain" in the IQs of culturally disadvantaged preschoolers on tests depending on specific information such as being able to name parts of the body and knowing names of familiar objects. Children who have not picked up this information at home get it quickly in nursery school and kindergarten.

In addition to these factors, something else operates to boost scores five to ten points from first to second test, provided the first test is really the first. When I worked in a psychological clinic, I had to give individual intelligence tests to a variety of children, a good many of whom came from an impoverished background. Usually I felt these children were really brighter than their IQ would indicate. They often appeared inhibited in their responsiveness in the testing situation on their first visit to my office, and when this was the case I usually had them come in on two to four different days for half-hour sessions with me in a "play therapy" room, in which we did nothing more than get better acquainted by playing ball, using finger paints, drawing on the blackboard, making things out of clay, and so forth. As soon as the child seemed to be completely at home in this setting, I would retest him on a parallel form of the Stanford-Binet. A boost in IQ of 8 to 10 points or so was the rule; it rarely failed, but neither was the gain very often much above this. So I am inclined to doubt that IQ gains up to this amount in young disadvantaged children have much of anything to do with changes in ability. They are largely a result simply of getting a more accurate IQ by testing under more optimal conditions. Part of creating more optimal conditions in the case of disadvantaged children consists of giving at least two tests, the first only for practice and for letting the child get to know the examiner. I would put very little confidence in a single test score, especially if it is the child's first test and more especially if the child is from a poor background and of a different race from the examiner. But I also believe it is possible to obtain accurate assessments of a child's ability, and I would urge that attempts to evaluate preschool enrichment programs measure the gains against initially valid scores. If there is not evidence that this precaution has been taken, and if there is no control group, one might as well subtract at least 5 points from the gain scores as having little or nothing to do with real intellectual growth.

It is interesting that the IQ gains typically found in enrichment programs are of about the same magnitude and durability as those found in studies of the effects of direct coaching and practice on intelligence tests. The average IQ gain in such studies is about nine or ten points (Vernon, 1954).

What Is Really Changed When We Boost IQ? Test scores may increase after special educational treatment, but one must then ask which components of test variance account for the gain. Is it *g* that gains, or is it something less central to our concept of intelligence? We will not know for sure until someone does a factor analysis of pre- and post-test scores, including a number of "reference" tests that were not a part of the pre-test battery. We should also factor analyze the tests at the item level, to see which types of test items reflect the most gain. Are they the items with the highest cultural loadings? It is worth noting that the studies showing authentic gains used tests which are relatively high in cultural loading. I have found no studies that demonstrated gains in relatively noncultural or nonverbal tests like Cattell's Culture Fair Tests and Raven's Progressive Matrices.

Furthermore, if gain consists of actual improvement in cognitive skills rather than of acquisition of simple information, it must be asked whether the gain in skill represents the intellectual skill that the test normally measures, and which, because of the test's high heritability, presumably reflects some important, biologically based aspect of mental development. Let me cite one example. In a well-known experiment Gates and Taylor (1925) gave young children daily practice over several months in repeating auditory digit series, just like the digit span subtests in the Wechsler and Stanford-Binet. The practice resulted in a marked gain in the children's digit span, equivalent to an IQ gain of about 20 points. But when the children were retested after an interval of six months without practicing digit recall, their digit performance was precisely at the level expected for their mental age as determined by other tests. The gains had been lost, and the digit test once again accurately reflected the children's overall level of mental development, as it did before the practice period. The well-known later "fading" of IQ gains acquired early in enrichment programs may be a similar phenomenon.

But there is another phenomenon that probably is even more important as one of the factors working against the persistence of initial gains. This is the so-called "cumulative deficit" phenomenon, the fact that many children called disadvantaged show a decline in IQ from preschool age through at least elementary

school. The term "cumulative deficit" may not be inappropriate in its connotations with respect to scholastic attainment, but it is probably a misleading misnomer when applied to the normal negatively accelerated growth rate of developmental characteristics such as intelligence. The same phenomenon can be seen in growth curves of stature, but no one would refer to the fact that some children gain height at a slower rate and level off at a lower asymptote as a "cumulative deficit." In short, it seems likely that some of the loss in initial gains is due to the more negatively accelerated growth curve for intelligence in disadvantaged children and is not necessarily due to waning or discontinuance of the instructional effort. The effort required to boost IQ from 80 to 90 at 4 or 5 years of age is miniscule compared to the effort that would be required by age 9 or 10. "Gains" for experimental children in this range, in fact, take the form of superiority over a control group which has declined in IQ; the "enriched" group is simply prevented from falling behind, so there is no absolute gain in IQ, but only an advantage relative to a declining control group. Because of the apparently ephemeral nature of the initial gains seen in preschool programs, judgments of these programs' effectiveness in making a significant impact on intellectual development should be based on long range results.

A further step in proving the effectiveness of a particular program is to demonstrate that it can be applied with comparable success by other individuals in other schools, and, if it is to be practicable on a large scale, to determine if it works in the hands of somewhat less inspired and less dedicated practitioners than the few who originated it or first put it into practice on a small scale. As an example of what can happen when a small-scale project gets translated to a large-scale one, we can note Kenneth B. Clark's (1963, p. 160) enthusiastic and optimistic description of a "total push" intensive compensatory program which originated in one school serving disadvantaged children in New York City, with initially encouraging results. Clark said, "These positive results can be duplicated in every school of this type." In fact, it was tried in 40 other New York schools, and became known as the Higher Horizons program. After three years of the program the children in it showed no gains whatever and even averaged slightly lower in achievement and IQ than similar children in ordinary schools (U.S. Commission on Civil Rights, 1967, p. 125).

Finally, little is known about the range of IQ most likely to show genuine gains under enrichment. None of the data I have seen in this area permits any clear judgment on this matter. It would be unwarranted to assume at this time that special educational programs push the whole IQ distribution up the scale,

77-943 O - 72 - 12

so that, for example, they would yield a higher precentage of children with IQs higher than two standard deviations above the mean. After a "total push" program, IQs, if they change at all, may no longer be normally distributed, so that the gains would not much affect the frequencies at the tails of the distribution. We simply do not know the answer to this at present, since the relevant data are lacking.

Hothouse or Fertilizer? There seems to be little doubt that a deprived environment can stunt intellectual development and that immersion in a good environment in early childhood can largely overcome the effects of deprivation, permitting the individual's genetic potential to be reflected in his performance. But can special enrichment and instructional procedures go beyond the prevention or amelioration of stunting? As Vandenberg (1968, p. 49) has asked, does enrichment act in a manner similar to a *hothouse,* forcing an early bloom which is nevertheless no different from a normal bloom, or does it act more like a *fertilizer,* producing bigger and better yields? There can be little question about the hothouse aspect of early stimulation and instruction. Within limits, children can learn many things at an earlier age than that at which they are normally taught in school. This is especially true of forms of associative learning which are mainly a function of time spent in the learning activity rather than of the development of more complex cognitive structures. While most children, for example, do not learn the alphabet until 5 or 6 years of age, they are fully capable of doing so at about 3, but it simply requires more time spent in learning. The cognitive structures involved are relatively simple as compared with, say, learning to copy a triangle or a diamond. Teaching a 3-year-old to copy a diamond is practically impossible; at five it is extremely difficult; at seven the child apparently needs no "teaching"—he copies the diamond easily. And the child of five who has been *taught* to copy the diamond seems to have learned something different from what the seven-year-old "knows" who can do it without being "taught." Though the final performance of the five-year-old and the seven-year-old may *look* alike, we know that the cognitive structures underlying their performance are different. Certain basic skills can be acquired either associatively by rote learning or cognitively by conceptual learning, and what superficially may appear to be the same performance may be acquired in preschoolers at an associative level, while at a conceptual level in older children. Both the four-year-old and the six-year-old may know that $2 + 2 = 4$, but this knowledge can be associative or cognitive. Insufficient attention has been given in preschool programs so far to the shift from associative to cognitive learning. The preschooler's capacity for associative learn-

ing is already quite well developed, but his cognitive or conceptual capacities are as yet rudimentary and will undergo their period of most rapid change between about five and seven years of age (White, 1965). We need to know more about what children can learn before age five that will transfer positively to later learning. Does learning something on an associative level facilitate or hinder learning the same content on a conceptual level?

While some preschool and compensatory programs have demonstrated earlier than normal learning of certain skills, the evidence for accelerating cognitive development or the speed of learning is practically nil. But usually this distinction is not made between sheer performance and the nature of the cognitive structures which support the gains in performance, and so the research leaves the issue in doubt. The answer to such questions is to be found in the study of the kinds and amount of transfer that result from some specific learning. The capacity for transfer of training is one of the essential aspects of what we mean by intelligence. The IQ gains reported in enrichment studies appear to be gains more in what Cattell calls 'crystallized," in contrast to "fluid," intelligence. This is not to say that gains of this type are not highly worthwhile. But having a clearer conception of just what the gains consist of will give us a better idea of how they can be most effectively followed up and of what can be expected of their effects on later learning and achievement.

Specific Programs. Hodges and Spicker (1967) have summarized a number of the more substantial preschool intervention studies designed to improve the intellectual capabilities and scholastic success of disadvantaged children. Here are some typical examples.

The *Indiana Project* focused on deprived Appalachian white children five years of age, with IQs in the range of 50 to 85. The children spent one year in a special kindergarten with a structured program designed to remedy specific diagnosed deficiencies of individual children in the areas of language development, fine motor coordination, concept formation, and socialization. Evaluation extended over two years, and gains were measured against three control groups: regular kindergarten, children who stayed at home during the kindergarten year, and children at home in another similar community. The average gain (measured against all three controls) after two years was 10.8 IQ points on the Stanford-Binet (final IQ 97.4) and 4.0 IQ points on the Peabody Picture Vocabulary Test (final IQ 90.4).

The *Perry Preschool Project* at Ypsilanti, Michigan, also was directed at disadvantaged preschool children with IQs between 50 and 85. The program was aimed at remedying lacks largely in the verbal prerequisites for first-grade learning and involved the parents as well as the children. There was a significant gain of 8.9 IQ points in the Stanford-Binet after one year of the preschool, but by the end of second grade the experimental group exceeded the controls, who had had no preschool attendance, by only 1.6 IQ points, a nonsignificant gain.

The *Early Training Project* under the direction of Gray and Klaus at Peabody College is described as a multiple intervention program, meaning that it included not only preschool enrichment but work with the disadvantaged children's mothers to increase their ability to stimulate their child's cognitive development at home. Two experimental groups, with two and three summers of preschool enrichment experience in a special school plus home visits by the training staff, experienced an average gain, four years after the start of the program, of 7.2 IQ points over a control group on the Stanford-Binet (final IQ of *E* group was 93.6).

The *Durham Education Improvement Program* (1966-1967b) has focused on preschool children from impoverished homes. The basic assumption of the program is stated as follows: "First, Durham's disadvantaged youngsters are considered normal at birth and potentially normal academic achievers, though they are frequently subjected to conditions jeopardizing their physical and emotional health. It is further assumed that they adapt to their environment according to the same laws of learning which apply to all children." The program is one of the most comprehensive and intensive efforts yet made to improve the educability of children from backgrounds of poverty. The IQ gains over about an eight to nine months' interval for various groups of preschoolers in the program are raw pre-post test gains, not gains over a control group. The average IQ gains on three different tests were 5.32 (Peabody Picture Vocabulary), 2.62 (Stanford-Binet), and 9.27 (Wechsler Intelligence Scale for children). In most cases, IQs changed from the 80s to the 90s.

The well-known Bereiter-Engelmann (1966) program at the University of Illinois is probably the most sharply focused of all. It aims not at all-round enrichment of the child's experience but at teaching specific cognitive skills, particularly of a logical, semantic nature (as contrasted with more diffuse "verbal stimulation"). The emphasis is on information processing skills considered essential for school learning. The Bereiter-Engelmann preschool is said to be academically oriented, since each day throughout the school year the children receive twenty-minute periods of intensive instruction in three major content areas—lan-

guage, reading, and arithmetic. The instruction, in small groups, explicitly involves maintaining a high level of attention, motivation, and participation from every child. Overt and emphatic repetition by the children are important ingredients of the instructional process. The pre-post gains (not measured against a control group) in Stanford-Binet IQ over an eighteen months' period are about 8 to 10 points. Larger gains are shown in tests that have clearly identifiable content which can reflect the areas receiving specific instruction, such as the Illinois Test of Psycholinguistic Abilities and tests of reading and arithmetic (Bereiter & Engelmann, 1968). The authors note that the gains are shared about equally by all children.

Bereiter and Engelmann, correctly, I believe, put less stock in the IQ gains than in the gains in scholastic performance achieved by the children in their program. They comment that the children's IQs were still remarkably low for children who performed at the academic level actually attained in the program. Their scholastic performance was commensurate with that of children 10 or 20 points higher in IQ. Such is the advantage of highly focused training—it can significantly boost the basic skills that count most. Bereiter and Engelmann (1966, p. 54) comment, "... to have taught children in a two-hour period per day enough over a broad area to bring the average IQ up to 110 or 120 would have been an impossibility." An important point of the Bereiter-Engelmann program is that it shows that scholastic performance—the acquisition of the basic skills—can be boosted much more, at least in the early years, than can the IQ, and that highly concentrated, direct instruction is more effective than more diffuse cultural enrichment.

The largest IQ gains I have seen and for which I was also able to examine the data and statistical analyses were reported by Karnes (1968), whose preschool program at the University of Illinois is based on an intensive attempt to ameliorate specific learning deficits in disadvantaged three-year-old children. Between the average age of 3 years 3 months and 4 years 1 month, children in the program showed a gain of 16.9 points in the Stanford-Binet IQ, while a control group showed a loss of 2.8 over the same period, making for a net gain of 19.7 IQ points for the experimental group. Despite rather small samples (E = 15, C = 14), this gain is highly significant statistically (a probability of less than 1 in 1000 of occurring by chance). Even so, I believe such findings need to be replicated for proper evaluation, and the durability of the gains needs to be assessed by follow-up studies over several years. There remains the question of the extent to which specific learning at age three affects cognitive structures which normally do not emerge

until six or seven years of age and whether induced gains at an early level of mental development show appreciable "transfer" to later stages. It is hoped that investigators can keep sufficient track of children in preschool programs to permit a later follow-up which could answer these questions. An initial small sample size mitigates against this possibility, and so proper research programs should be planned accordingly.

"Expectancy Gain." Do disadvantaged children perform relatively poorly on intelligence tests because their teachers have low expectations for their ability? This belief has gained popular currency through an experiment by Rosenthal and Jacobson (1968). Their notion is that the teacher's expectations for the child's performance act as a self-fulfilling prophecy. Consequently, according to this hypothesis, one way to boost these children's intelligence, and presumably their general scholastic performance as well, is to cause teachers to hold out higher expectations of these children's ability. To test this idea, Rosenthal and Jacobson picked about five children at random from each of the classes in an elementary school and then informed the classroom teachers that, according to test results, the selected children were expected to show unusual intellectual gains in the coming year. Since the "high expectancy" children in each class were actually selected at random, the only way they differed from their classmates was presumably in the minds of their teachers. Group IQ tests administered by the teachers on three occasions during the school year showed a significantly larger gain in the "high expectancy" children than in their classmates. Both groups gained in IQ by amounts that are typically found as a result of direct coaching or of "total push" educational programs. Yet the authors note that "Nothing was done directly for the disadvantaged child at Oak School. There was no crash program to improve his reading ability, no special lesson plans, no extra time for tutoring, no trips to museums or art galleries. There was only the belief that the children bore watching, that they had intellectual competencies that would in due course be revealed" (p. 181). The net total IQ gain (i.e., Expectancy group minus Control group) for all grades was 3.8 points. Net gain in verbal IQ was 2.1; for Reasoning (nonverbal) IQ the gain was 7.2. Differences were largest in grades 1 and 2 and became negligible in higher grades. The statistical significance of the gains is open to question and permits no clear-cut conclusion. (The estimation of the error variance is at issue: the investigators emphasized the individual pupil's scores as the unit of analysis rather than the means of the E and C groups for each classroom as the unit. The latter procedure, which is regarded as more rigorous by many statisticians, yields statistically negligible results.)

Because of the questionable statistical significance of the results of this study, there may actually be no phenomenon that needs to be explained. Other questionable aspects of the conduct of the experiment make it mandatory that its results be replicated under better conditions before any conclusions from the study be taken seriously or used as a basis for educational policy. For example, the same form of the group-administered IQ test was used for each testing, so that specific practice gains were maximized. The teachers themselves administered the tests, which is a faux pas par excellence in research of this type. The dependability of teacher-administered group tests leaves much to be desired. Would any gains beyond those normally expected from general test familiarity have been found if the children's IQs had been accurately measured in the first place by individual tests administered by qualified psychometrists without knowledge of the purpose of the experiment? These are some of the conditions under which such an experiment must be conducted if it is to inspire any confidence in its results.

Conclusions About IQ Gains. The evidence so far suggests the tentative conclusion that the pay-off of preschool and compensatory programs in terms of IQ gains is small. Greater gains are possible in scholastic performance when instructional techniques are intensive and highly focused, as in the Bereiter-Engelmann program. Educators would probably do better to concern themselves with teaching basic skills directly than with attempting to boost overall cognitive development. By the same token, they should deemphasize IQ tests as a means of assessing gains, and use mainly direct tests of the skills the instructional program is intended to inculcate. The techniques for raising intelligence per se, in the sense of *g*, probably lie more in the province of the biological sciences than in psychology and education.

Gordon and Wilkerson (1966, pp. 158-159) have made what seems to me perhaps the wisest statement I have encountered regarding the proper aims of intervention programs:

> . . . the unexpressed purpose of most compensatory programs is to make disadvantaged children as much as possible like the kinds of children with whom the school has been successful, and our standard of educational success is how well they approximate middle-class children in school performance. It is not at all clear that the concept of compensatory education is the one which will most appropriately meet the problems of the disadvantaged. These children are *not* middle-class children, many of them never *will* be, and they can never be anything but second-rate as long as they are thought of as po-

tentially middle-class children.... At best they are different, and an approach which views this difference merely as something to be overcome is probably doomed to failure.

"Learning Quotient" versus Intelligence Quotient

If many of the children called culturally disadvantaged are indeed "different" in ways that have educational implications, we must learn as much as possible about the real nature of these differences. To what extent do the differences consist of more than just the well-known differences in IQ and scholastic achievement, and, of course, the obvious differences in cultural advantages in the home?

Evidence is now emerging that there are stable ethnic differences in *patterns* of ability and that these patterns are invariant across wide socioeconomic differences (Lesser, Fifer, & Clark, 1965; Stodolsky & Lesser, 1967). Middle-class and lower-class groups differed about one standard deviation on all four abilities (Verbal, Reasoning, Number, Space) measured by Lesser and his co-workers, but the profile or pattern of scores was distinctively different for Chinese, Jewish, Negro, and Puerto Rican children, regardless of their social class. Such differences in patterns of ability are bound to interact with school instruction. The important question is how many other abilities there are that are not tapped by conventional tests for which there exist individual and group differences that interact with methods of instruction.

Through our research in Berkeley we are beginning to perceive what seems to be a very significant set of relationships with respect to patterns of ability which, unlike those of Lesser et al., seem to interact more with social class than with ethnic background.

In brief, we are finding that a unidimensional concept of intelligence is quite inadequate as a basis for understanding social class differences in ability. For example, the magnitude of test score differences between lower- and middle-class children does not always correspond to the apparent "cultural loading" of the test. Some of the least culturally loaded tests show the largest differences between lower- and middle-class children. At least two dimensions must be postulated to comprehend the SES differences reported in the literature and found in our laboratory (see Jensen, 1968c, 1968d). These two dimensions and the hypothetical location of various test loadings on each dimension are shown in Figure 17. The horizontal axis represents the degree of cultural loading of the test. It is defined by the test's heritability. I have argued elsewhere (Jensen, 1968c) that the heritability index for a test is probably our best objective criterion of its culture-

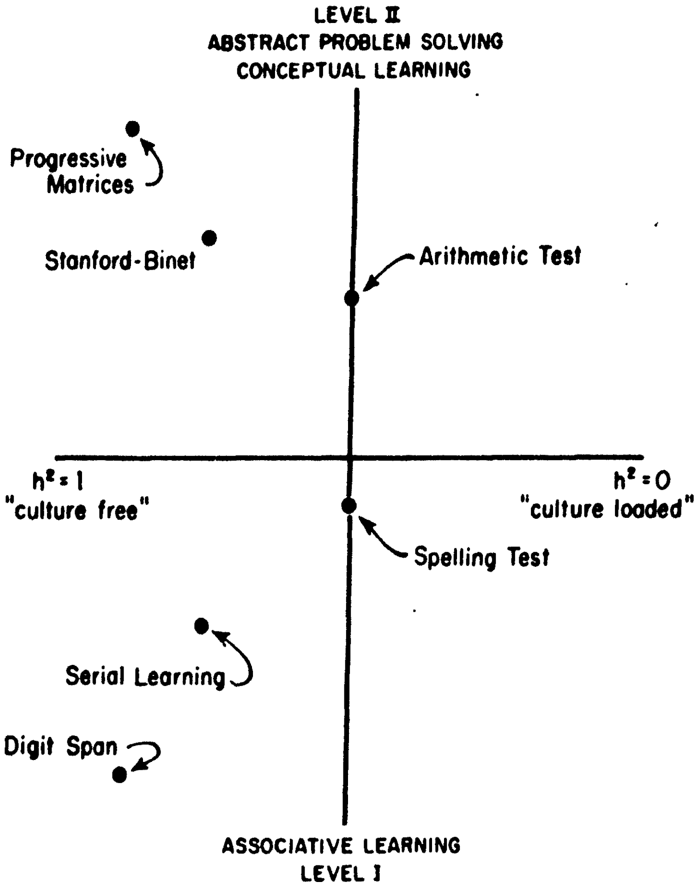


FIGURE 17.
The two-dimensional space required for comprehending social class differences in performance on tests of intelligence, learning ability, and scholastic achievement. The locations of the various "tests" are hypothetical.

fairness. Just because tests do not stand at one or the other extreme of this continuum does not mean that the concept of culture-fairness is not useful in discussing psychological tests. The vertical axis in Figure 17 represents a continuum ranging from "simple" associative learning to complex cognitive or conceptual learning. I have hypothesized two genotypically distinct basic processes underlying this continuum, labeled Level I (associative ability) and Level II (conceptual ability). Level I involves the neural registration and consolidation of stimulus

inputs and the formation of associations. There is relatively little transformation of the input, so there is a high correspondence between the forms of the stimulus input and the form of the response output. Level I ability is tapped mostly by tests such as digit memory, serial rote learning, selective trial-and-error learning with reinforcement (feedback) for correct responses, and in slightly less "pure" form by free recall of visually or verbally presented materials, and paired-associate learning. Level II abilities, on the other hand, involve self-initiated elaboration and transformation of the stimulus input before it eventuates in an overt response. Concept learning and problem solving are good examples. The subject must actively manipulate the input to arrive at the output. This ability is best measured by intelligence tests with a low cultural loading and a high loading on g—for example, Raven's Progressive Matrices.

Social class differences in test performance are more strongly associated with the vertical dimension in Figure 17 than with the horizontal.

Associative Learning Ability

Teachers of the disadvantaged have often remarked that many of these children seem much brighter than their IQs would lead one to expect, and that, even though their scholastic performance is usually as poor as that of middle-class children of similar IQ, the disadvantaged children usually appear much brighter in nonscholastic ways than do their middle-class counterparts in IQ. A lower-class child coming into a new class, for example, will learn the names of 20 or 30 children in a few days, will quickly pick up the rules and the know-how of various games on the playground, and so on—a kind of performance that would seem to belie his IQ, which may even be as low as 60. This gives the impression that the test is "unfair" to the disadvantaged child, since middle-class children in this range of IQ will spend a year in a classroom without learning the names of more than a few classmates, and they seem almost as inept on the playground and in social interaction as they are in their academic work.

We have objectified this observation by devising tests which can reveal these differences. The tests measure associative learning ability and show how fast a child can learn something relatively new and unfamiliar, right in the test situation. The child's performance does not depend primarily, as it would in conventional IQ tests, upon what he has already learned at home or elsewhere before he comes to take the test. We simply give him something to learn, under conditions which permit us to measure the rate and thoroughness of the learning. The tasks most frequently used are various forms of auditory digit memory, learning

the serial order of a number of familiar objects or pictures of objects, learning to associate pairs of pictures of familiar objects, and free recall of names or objects presented from one to five times in a random order.

Our findings with these tests, which have been presented in greater detail elsewhere (Jensen, 1968a, 1968b, 1968d, 1968e; Jensen, 1968f; Jensen & Rohwer, 1968), seem to me to be of great potential importance to the education of many of the children called disadvantaged. What we are finding, briefly, is this: lower-class children, whether white, Negro, or Mexican-American, perform as well on these direct learning tests as do middle-class children. Lower-class children in the IQ range of about 60 to 80 do markedly *better* than middle-class children who are in this range of IQ. Above about IQ 100, on the other hand, there is little or no difference between social class groups on the learning tests.

At first we thought we had finally discovered a measure of "culture-fair" testing, since we found no significant SES differences on these learning tests. But we can no longer reconcile this interpretation with all the facts now available. Some of the low SES children with low IQs on culturally loaded tests, like the Peabody Picture Vocabulary Tests, do very well on our learning tests, but do not have higher IQs on less culturally loaded tests of *g*, like the Progressive Matrices. It appears that we are dealing here with two kinds of abilities—associative learning ability (Level I) and cognitive or conceptual learning and problem-solving ability (Level II).

One particular test—free recall—shows the distinction quite well, since a slight variation in the test procedure makes the difference between whether it measures Level I or Level II. This is important, because it is sometimes claimed that low SES children do better on our learning tests than on IQ tests because the former are more interesting or more "relevant" to them, and thus make them more highly motivated to perform at their best. This is not a valid interpretation, since when essentially the same task is made either "associative" or "cognitive," we get differences of about one standard deviation in the mean scores of lower- and middle-class children. For example, 20 unrelated familiar objects (doll, toy car, comb, cup, etc.) are shown to children who are then asked to recall as many objects as they can in any order that may come to mind. The random presentation and recall are repeated five times to obtain a more reliable score. Lower- and middle-class elementary school children perform about the same on this task, although they differ some 15 to 20 points in IQ. This free recall test has a low correlation with IQ and the correlation is lower for the low SES children. But then we can change the recall test so that it gives quite different results.

This is shown in an experiment from our laboratory by Glasman (1968). (In this study SES and race are confounded, since the low SES group were Negro children and the middle SES group were white.) Again, 20 familiar objects are presented, but this time the objects are selected so that they can be classified into one of four categories, *animals, furniture, clothing,* or *foods.* There are five items in each of the four categories, but all 20 items are presented in a random order on each trial. Under this condition a large social class difference shows up: the low SES children perform only slightly better on the average than they did on the uncategorized objects, while the middle SES children show a great improvement in performance which puts their scores about one standard deviation above the low SES children. Furthermore, there is much greater evidence of "clustering" the items in free recall for the middle SES than for the low SES children. That is, the middle-class children rearrange the input in such a way that the order of output in recall corresponds to the categories to which the objects may be assigned. The low SES children show less clustering in this fashion, although many show rather idiosyncratic pair-wise "clusters" that persist from trial to trial. There is a high correlation between the strength of the clustering tendency and the amount of recall. Also, clustering tendency is strongly related to age. Kindergarteners, for example, show little difference between recall of categorized and uncategorized lists, and at this age SES differences in performance are nil. By fourth or fifth grade, however, the SES differences in clustering tendency are great, with a correspondingly large difference in ability to recall categorized lists.

It is interesting, also, that the recall of categorized lists correlates highly with IQ. In fact, when mental age or IQ is partialled out of the results, there are no significant remaining SES differences in recall. Post-test interviews showed that the recall differences for the two social class groups cannot be attributed to the low SES group's not knowing the category names. The children know the categories but tend not to use them spontaneously in recalling the list.

In general, we find that Level I associative learning tasks correlate very substantially with IQ among middle-class children but have very low correlations with IQ among lower-class children (Jensen, 1968b). The reason for this difference in correlations can be traced back to the form of the scatter diagrams for the middle and low SES groups, which is shown schematically in Figure 18. Since large representative samples of the entire school population have not been studied so far, the exact form of the correlation scatter diagram has not yet been well established, but the schematic portrayal of Figure 18 is what could be most reasonably hypothesized on the basis of several lines of evidence now available. (Data

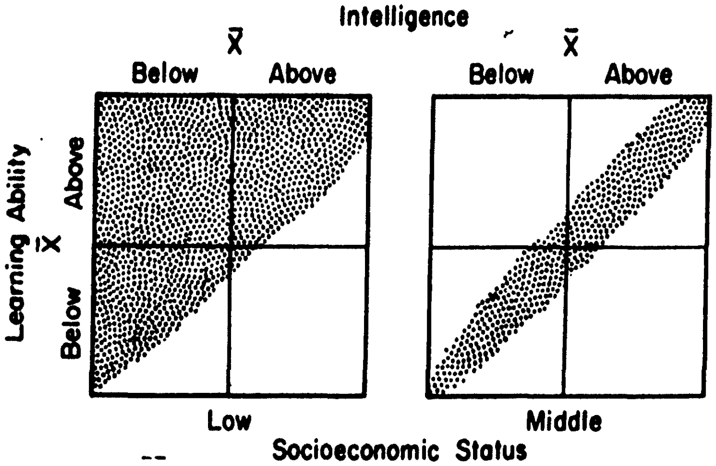


FIGURE 19.

Schematic illustration of the essential form of the correlation scatter-diagram for the relationship between associative learning ability and IQ in Low SES and Upper-Middle SES groups.

on a representative sample of 5000 children given Level I and Level II tests are now being analyzed to establish the forms of the correlation plots for low and middle SES groups.) The form of the correlation as it now appears suggests a hierarchical arrangement of mental abilities, such that Level I ability is necessary but not sufficient for Level II. That is, high performance on Level II tasks depends upon better than average ability on Level I, but the reverse does not hold. If this is true, the data can be understood in terms of one additional hypothesis, namely, that Level I ability is distributed about the same in all social class groups, while Level II ability is distributed differently in lower and middle SES groups. The hypothesis is expressed graphically in Figure 19. Heritability studies of Level II tests cause me to believe that Level II processes are not just the result of interaction between Level I learning ability and experientially acquired strategies or learning sets. That learning is necessary for Level II no one doubts, but certain neural structures must also be available for Level II abilities to develop, and these are conceived of as being different from the neural structures underlying Level I. The genetic factors involved in each of these types of ability are presumed to have become differentially distributed in the population as a function of social class, since Level II has been most important for scholastic performance under the traditional methods of instruction.

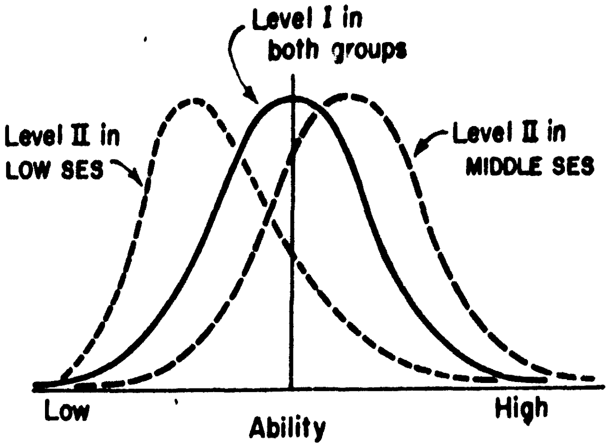


FIGURE 18.
Hypothetical distributions of Level I (solid line) and Level II (dashed line) abilities in middle-class and culturally disadvantaged populations.

From evidence on age differences in different tasks on the Level I—Level II continuum (e.g., Jensen & Rohwer, 1965), I have suggested one additional hypothesis concerning the developmental rates of Level I and Level II abilities in lower and middle SES groups, as depicted in Figure 20. Level I abilities are seen as developing rapidly and as having about the same course of development and final level in both lower and middle SES groups. Level II abilities, by contrast, develop slowly at first, attain prominence between four and six years of age, and show an increasing difference between the SES groups with increasing age. This formulation is consistent with the increasing SES differences in mental age on standard IQ tests, which tap mostly Level II ability.

Thus, ordinary IQ tests are not seen as being "unfair" in the sense of yielding inaccurate or invalid measures for the many disadvantaged children who obtain low scores. If they are unfair, it is because they tap only one part of the total spectrum of mental abilities and do not reveal that aspect of mental ability which may be the disadvantaged child's strongest point—the ability for associative learning.

Since traditional methods of classroom instruction were evolved in populations having a predominantly middle-class pattern of abilities, they put great emphasis

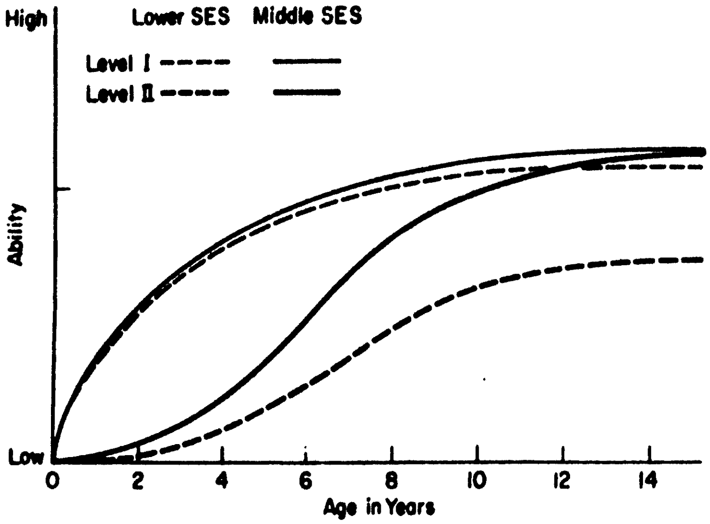


FIGURE 20.
Hypothetical growth curves for Level I and Level II abilities in middle SES and low SES populations.

on cognitive learning rather than associative learning. And in the post-Sputnik era, education has seen an increased emphasis on cognitive and conceptual learning, much to the disadvantage of many children whose mode of learning is predominantly associative. Many of the basic skills can be learned by various means, and an educational system that puts inordinate emphasis on only one mode or style of learning will obtain meager results from the children who do not fit this pattern. At present, I believe that the educational system—even as it falteringly attempts to help the disadvantaged—operates in such a way as to maximize the importance of Level II (i.e., intelligence or *g*) as a source of variance in scholastic performance. Too often, if a child does not learn the school subject matter when taught in a way that depends largely on being average or above average on *g*, he does not learn at all, so that we find high school students who have failed to learn basic skills which they could easily have learned many years earlier by means that do not depend much on *g*. It may well be true that many children today are confronted in our schools with an educational philosophy and methodology which were mainly shaped in the past, entirely without any roots in these children's genetic and cultural heritage. The educational system was never

allowed to evolve in such a way as to maximize the actual potential for learning that is latent in these children's patterns of abilities. If a child cannot show that he "understands" the meaning of $1 + 1 = 2$ in some abstract, verbal, cognitive sense, he is, in effect, not allowed to go on to learn $2 + 2 = 4$. I am reasonably convinced that all the basic scholastic skills can be learned by children with normal Level I learning ability, provided the instructional techniques do not make *g* (i.e., Level II) the *sine qua non* of being able to learn. Educational researchers must discover and devise teaching methods that capitalize on existing abilities for the acquisition of those basic skills which students will need in order to get good jobs when they leave school. I believe there will be greater rewards for all concerned if we further explore different types of abilities and modes of learning, and seek to discover how these various abilities can serve the aims of education. This seems more promising than acting as though only one pattern of abilities, emphasizing *g*, can succeed educationally, and therefore trying to inculcate this one ability pattern in all children.

If the theories I have briefly outlined here become fully substantiated, the next step will be to develop the techniques by which school learning can be most effectively achieved in accordance with different patterns of ability. By all means, schools must discover *g* wherever it exists and see to it that its educational correlates are fully encouraged and cultivated. There can be little doubt that certain educational and occupational attainments depend more upon *g* than upon any other single ability. But schools must also be able to find ways of utilizing other strengths in children whose major strength is not of the cognitive variety. One of the great and relatively untapped reservoirs of mental ability in the disadvantaged, it appears from our research, is the basic ability to learn. We can do more to marshal this strength for educational purposes.

If diversity of mental abilities, as of most other human characteristics, is a basic fact of nature, as the evidence indicates, and if the ideal of universal education is to be successfully pursued, it seems a reasonable conclusion that schools and society must provide a range and diversity of educational methods, programs, and goals, and of occupational opportunities, just as wide as the range of human abilities. Accordingly, the ideal of equality of educational opportunity should not be interpreted as uniformity of facilities, instructional techniques, and educational aims for all children. Diversity rather than uniformity of approaches and aims would seem to be the key to making education rewarding for children of different patterns of ability. The reality of individual differences thus need not mean educational rewards for some children and frustration and defeat for others.

References

Altus, W. D. Birth order and its sequelae. *Science*, 1966, 151, 44-59.
Anastasi, A. Intelligence and family size. *Psychol. Bull.*, 1956, 53, 187-209.
Bajema, C. J. Estimation of the direction and intensity of natural selection in relation to human intelligence by means of the intrinsic rate of natural increase. *Eugen. Quart.*, 1963, 10, 175-187.
Bajema, C. J. Relation of fertility to educational attainment in a Kalamazoo public school population: A follow-up study. *Eugen. Quart.*, 1966, 13, 306-315.
Bayley, N. Research in child development: A longitudinal perspective. *Merrill-Palmer Quart. Behav. Developm.*, 1965, 11, 183-208. (b)
Bayley, N. Comparisons of mental and motor test scores for ages 1-15 months by sex, birth order, race, geographical location, and education of parents. *Child Developm.*, 1965, 36, 379-411. (a)
Bayley, N. Learning in adulthood: The role of intelligence. In H. J. Klausmeier & C. W. Harris (Eds.), *Analyses of concept learning*. New York: Academic Press, 1966. Pp. 117-138.
Bayley, N. Behavioral correlates of mental growth: Birth to thirty-six years. *Amer. Psychol.*, 1968, 23, 1-17.
Bereiter, C., & Engelmann, S. *Teaching disadvantaged children in the preschool*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.
Bereiter, C., & Engelmann, S. An academically oriented preschool for disadvantaged children: Results from the initial experimental group. In D. W. Brison & J. Hill (Eds.), *Psychology and early childhood education*. Ontario Institute for Studies in Education, 1968. No. 4. Pp. 17-36.
Bilodeau, E. A. (Ed.), *Acquisition of skill*. New York: Academic Press, 1966.
Brison, D. W. Can and should learning be accelerated? In D. W. Brison (Ed.), *Accelerated learning and fostering creativity*. Toronto, Canada: Ontario Institute for Studies in Education, 1968. Pp. 5-9.
Bronfenbrenner, U. The psychological costs of quality and equality in education. *Child Developm.*, 1967, 38, 909-925.
Buck, C. Discussion of "Culturally related reproductive factors in mental retardation" by Graves et al. Paper read at Conference on Sociocultural Aspects of Mental Retardation, Peabody College, Nashville, Tenn., June, 1968.
Burks, B. S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. *Yearb. Nat. Soc. Stud. Educ.*, 1928, 27, (I), 219-316.
Burt, C. The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, 1955, 25, 158-177.
Burt, C. The distribution of intelligence. *Brit. J. Psychol.*, 1957, 48, 161-175.
Burt, C. The inheritance of mental ability. *Amer. Psychol.*, 1958, 13, 1-15.
Burt, C. Class difference in general intelligence: III. *Brit. J. Stat. Psychol.*, 1959, 12, 15-33.
Burt, C. Intelligence and social mobility. *Brit. J. Stat. Psychol.*, 1961, 14, 3-24.
Burt, C. Is intelligence distributed normally? *Brit. J. Stat. Psychol.*, 1963, 16, 175-190.
Burt, C. The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *Brit. J. Psychol.*, 1966, 57, 137-153.
Burt, C. Mental capacity and its critics. *Bull. Brit. Psychol. Soc.*, 1968, 21, 11-18.
Burt, C., & Howard, M. The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Stat. Psychol.*, 1956, 9, 95-131.
Burt, C., & Howard, M. The relative influence of heredity and environment on assessments of intelligence. *Brit. J. Stat. Psychol.*, 1957, 10, 99-104.
Bloom, B. S. *Stability and change in human characteristics*. New York: Wiley, 1964.
Carter, C. O. Differential fertility by intelligence. In J. E. Meade & A. S. Parkes (Eds.), *Genetic and environmental factors in human ability*. New York: Plenum Press, 1966. Pp. 185-200.
Cattell, R. B. The multiple abstract variance analysis equations and solutions: For nature-nurture research on continuous variables. *Psychol. Rev.*, 1960, 67, 353-372.

77-942 O - 72 - 13

Cattell, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *J. educ. Psychol.*, 1963, 54, 1-22.

Churchill, J. A., Neff, J. W., & Caldwell, D. F. Birth weight and intelligence. *Obstetrics and gynecology*, 1966, 28, 425-429.

Clark, K. B. Educational stimulation of racially disadvantaged children. In A. H. Passow (Ed.), *Education in depressed areas*. New York: Teachers College Press, Columbia Univer., 1963. Pp. 142-162.

Coleman, J. S., et al. *Equality of educational opportunity*. U.S. Dept. of Health, Education, & Welfare, 1966.

Cooper, R., & Zubek, J. Effects of enriched and restricted early environments on the learning ability of bright and dull rats. *Canad. J. Psychol.*, 1958, 12, 159-164.

Cravioto, J. Malnutrition and behavioral development in the preschool child. *Pre-school child malnutrition*. National Health Science, Public., 1966, No. 1282.

Cravioto, J., De Licardie, E. R., & Birch, H. G. Nutrition, growth, and neurointegrative development: An experimental and ecologic study. *Pediatrics*, 1966, 38, 319-372.

Davis, K. Final note on a case of extreme isolation. *Amer. J. Sociol.*, 1947, 57, 432-457.

Deutsch, M., Katz, I., & Jensen, A. R. (Eds.) *Social class, race, and psychological development*. New York: Holt, Rinehart & Winston, 1968.

Dobzhansky, T. Genetic differences between people cannot be ignored. *Scientific Res.*, 1968, 3, 32-33. (a)

Dobzhansky, T. On genetics, sociology, and politics. *Perspect. Biol. Med.*, 1968, 11, 544-554. (b)

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. *Psychol. Bull.*, 1960, 57, 361-402.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States: 1959-1965. *Psychol. Bull.*, 1968 (*Monogr. Suppl.* 70, No. 3, Part 2).

Duncan, O. D. Is the intelligence of the general population declining? *Amer. sociol. Rev.*, 1952, 17, 401-407.

Duncan, O. D., Featherman, D. L., & Duncan, B. Socioeconomic background and occupational achievement: Extensions of a basic model. Final Report, Project No. 5-0074 (EO-191) U.S. Dept. of Health, Education, and Welfare, Office of Education, Bureau of Research, May, 1968.

Durham Education Improvement Program, 1966-1967. (a)

Durham Education Improvement Program, Research, 1966-1967. (b)

Dustman, R. E., & Beck, E. C. The visually evoked potential in twins. *Electroenceph. clin. Neurophysiol.*, 1965, 19, 570-575.

Eckland, B. K. Genetics and sociology: A reconsideration. *Amer. Soc. Rev.*, 1967, 32, 173-194.

Eells, K., et al. *Intelligence and cultural differences*. Chicago: Univ. Chicago Press, 1951.

Erlenmeyer-Kimling, L., & Jarvik, L. F. Genetics and intelligence: a review. *Science*, 1963, 142, 1477-1479.

Falconer, D. S. *An introduction to quantitative genetics*. New York: Ronald Press, 1960.

Freeman, R. A. Schools and the elusive 'average children' concept. *Wall Street Journal*, July 8, 1968, p. 12.

Fuller, J. L., & Thompson, W. R. *Behavior genetics*. New York: Wiley, 1960.

Gates, A. I., & Taylor, G. A. An experimental study of the nature of improvement resulting from practice in mental function. *J. educ. Psychol.*, 1925, 16, 583-593.

Geber, M. The psycho-motor development of African children in the first year, and the influence of maternal behavior. *J. soc. Psychol.*, 1958, 47, 185-195.

Geber, M., & Dean, R. F. A. The state of development of newborn African children. *Lancet*, 1957, 1216-1219.

Ghiselli, E. E. The measurement of occupational aptitude. *Univer. of California Publ. in Psychol.*, Vol. 8, No. 2. Berkeley, Calif.: Univer. of California Press, 1955.

Glasman, L. D. A social-class comparison of conceptual processes in children's free recall. Unpublished doctoral dissertation, Univer. of California, 1968.

Goertzel, V., & Goertzel, M. G. *Cradles of eminence*. London: Constable, 1962.

Goodenough, F. L. New evidence on environmental influence on intelligence. *Yearb. Nat. Soc. Stud. Educ.*, 1940, 39, Part I, 307-365.

Gordon, E. W., & Wilkerson, D. A. *Compensatory education for the disadvantaged.* New York: College Entrance Examination Board, 1966.

Gottesman, I. Genetic aspects of intelligent behavior. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 253-296.

Gottesman, I. Biogenetics of race and class. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.), *Social class, race, and psychological development.* New York: Holt, Rinehart & Winston, 1968. Pp. 11-51.

Graves, W. L., Freeman, M. G., & Thompson, J. D. Culturally related reproductive factors in mental retardation. Paper read at Conference on Sociocultural Aspects of Mental Retardation, Peabody College, Nashville, Tenn., June, 1968.

Guilford, J. P. *The nature of human intelligence.* New York: McGraw-Hill, 1967.

Hardy, J. B. Perinatal factors and intelligence. In S. F. Osler & R. E. Cooke (Eds.), *The biosocial basis of mental retardation.* Baltimore, Md.: The Johns Hopkins Press, 1965. Pp. 35-60.

Harlow, H. F., & Griffin, G. Induced mental and social deficits in Rhesus monkeys. In S. F. Osler & R. E. Cooke (Eds.), *The biosocial basis of mental retardation.* Baltimore, Md.: The Johns Hopkins Press, 1965. Pp. 87-106.

Harlow, H. F., & Harlow, M. K. The mind of man. In *Yearbook of science and technology.* New York: McGraw-Hill, 1962.

Harrell, R. F., Woodyard, E., & Gates, A. I. *The effects of mothers' diets on the intelligence of offspring.* New York: Bureau of Publications, Teachers College, 1955.

Heber, R. Research on education and habilitation of the mentally retarded. Paper read at Conference on Sociocultural Aspects of Mental Retardation, Peabody College, Nashville, Tenn., June, 1968.

Heber, R., Dever, R., & Conry, J. The influence of environmental and genetic variables on intellectual development. In H. J. Prehm, L. A. Hamerlynck, & J. E. Crosson (Eds.), *Behavioral Research in mental retardation.* Eugene, Oregon: Univer. of Oregon Press, 1968. Pp. 1-23.

Heyns, O. S. *Abdominal decompression.* Johannesburg: Witwatersrand Univer. Press, 1963.

Higgins, J., Reed, S., & Reed, E. Intelligence and family size: A paradox resolved. *Eugen. Quart.*, 1962, 9, 84-90.

Hill, A. C., & Jaffee, F. S. Negro fertility and family size preferences. In T. Parsons & K. B. Clark (Eds.), *The Negro American.* Cambridge, Mass.: Houghton-Mifflin, 1966. Pp. 134-159.

Hodges, W. L., & Spicker, H. H. The effects of preschool experiences on culturally deprived children. In W. W. Hartup & N. L. Smothergill (Eds.), *The young child: Reviews of research.* Washington, D.C.: National Association for the Education of Young Children, 1967. Pp. 262-289.

Honzik, M. P. Developmental studies of parent-child resemblance in intelligence. *Child Developm.*, 1957, 28, 215-228.

Honzik, M. P. The mental and motor test performances of infants diagnosed or suspected of brain injury. Final Rep., Contract SA 43 PH 2426. Washington, D.C.: National Institute of Health, National Institute of Neurological Diseases and Blindness, Collaborative Research, May, 1962.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Huntley, R. M. C. Heritability of intelligence. In J. E. Meade & A. S. Parker (Eds.), *Genetic and environmental factors in human ability.* New York: Plenum Press, 1966. Pp. 201-218.

Husén, T. Abilities of twins. *Scand. J. Psychol.*, 1960, 1, 125-135.

Jensen, A. R. Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proc. Nat. Acad. Sci.*, 1967, 58, 149-157. (a)

Jensen, A. R. Varieties of individual differences in learning. In Gagné, R. M. (Ed.), *Learning and individual differences.* Columbus, Ohio: Merrill, 1967. (b)

Jensen, A. R. Social class, race, and genetics: Implications for education. *Amer. Educ. Res. J.*, 1968, 5, 1-42. (a)

Jensen, A. R. Patterns of mental ability and socioeconomic status. *Proc. Nat. Acad. Sci.*, 1968, **60**, 1330-1337. (b)

Jensen, A. R. Another look at culture-fair tests. In *Western Regional Conference on Testing Problems, Proceedings for 1968*, "Measurement for Educational Planning." Berkeley, Calif.: Educational Testing Service, Western Office, 1968. Pp. 50-104. (c)

Jensen, A. R. Intelligence, learning ability, and socioeconomic status. *J. spec. Educ.*, 1968. (d)

Jensen, A. R. Social class and verbal learning. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.), *Social class, race, and psychological development*. New York: Holt, Rinehart & Winston, 1968. Pp. 115-174. (e)

Jensen, A. R. The culturally disadvantaged and the heredity-environment uncertainty. In J. Helmuth (Ed.), *The culturally disadvantaged child*. Vol. 2. Seattle, Wash.: Special Child Publications, 1968. (f)

Jensen, A. R., & Rohwer, W. D., Jr. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Developm.*, 1965, **36**, 601-608.

Jensen, A. R., & Rohwer, W. D., Jr. Mental retardation, mental age, and learning rate. *J. educ. Psychol.*, 1968.

Jepsen, N. P., & Bredmose, G. V. Investigations into the age of mentally deficient women at their first delivery. *Acta Psychiat. Scand.*, 1956 (Monogr. Suppl. 108), Pp. 203-210.

Jones, H. E. The environment and mental development. In L. Carmichael (Ed.), *Manual of child psychology*. (2nd ed.) New York: Wiley, 1954. Pp. 631-696.

Karnes, M. B. A research program to determine the effects of various preschool intervention programs on the development of disadvantaged children and the strategic age for such intervention. Paper read at Amer. Educ. Res. Ass., Chicago, Feb., 1968.

Kempthorne, O. *An introduction to genetic statistics*. New York: Wiley, 1957.

Kennedy, W. A., Van De Riet, V., & White, J. C., Jr. A normative sample of intelligence and achievement of Negro elementary school children in the Southeastern United States. *Monogr. Soc. Res. Child Developm.*, 1963, **28**, No. 6.

Kushlick, A. Assessing the size of the problem of subnormality. In J. E. Meade & A. S. Parkes (Eds.), *Genetic and environmental factors in human ability*. New York: Plenum Press, 1966. Pp. 121-147.

Kuttner, R. E. *Biochemical anthropology*. In R. E. Kuttner (Ed.), *Race and modern science*. New York: Social Science Press, 1967. Pp. 197-222.

Kuttner, R. E. Letters to and from the editor. *Perspect. Biol. Med.*, 1968, **11**, 707-709.

Lawrence, E. M. An investigation into the relation between intelligence and inheritance. *Brit. J. Psychol. Monogr. Suppl.*, 1931, **16**, No. 5.

Leahy, A. M. Nature-nurture and intelligence. *Genet. Psychol. Monogr.*, 1935, **17**, 241-305.

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. *Monogr. Soc. Res. Child Developm.*, 1965, **30**, (4).

Lindzey, G. Some remarks concerning incest, the incest taboo, and psychoanalytic theory. *Amer. Psychol.*, 1967, **22**, 1051-1059.

Loehlin, J. C. Psychological genetics, from the study of human behavior. In R. B. Cattell (Ed.), *Handbook of modern personality theory*. New York: Aldine, in press.

Loevinger, J. On the proportional contributions of differences in nature and nurture to differences in intelligence. *Psychol. Bull.*, 1943, **40**, 725-756.

Medical World News. Using speed of brain waves to test IQ. 1968, **9**, 26.

Mitra, S. Income, socioeconomic status, and fertility in the United States. *Eugen. Quart.*, 1966, **13**, 223-230.

Money, J. Two cytogenetic syndromes: Psychologic comparisons 1. Intelligence and specific-factor quotients. *J. psychiat. Res.*, 1964, **2**, 223-231.

Moynihan, D. P. *The Negro family*. Washington, D.C.: Office of Policy Planning and Research, United States Department of Labor, 1965.

Moynihan, D. P. Employment, income, and the ordeal of the Negro family. In T. Parsons & K. B. Clark (Eds.), *The Negro American*. Cambridge, Mass.: Houghton-Mifflin, 1966. Pp. 134-159.

National Academy of Sciences. Racial studies: Academy states position on call for new research. *Science*, 1967, 158, 892-893.
Naylor, A. E., & Myrianthopoulos, N. C. The relation of ethnic and selected socioeconomic factors to human birth-weight. *Ann. Hum. Genet.*, 1967, 31, 71-83.
Nelson, G. K., & Dean, R. F. A. *Bull. World Health Organ.*, 1959, 21, 779. Cited by G. Cravioto, Malnutrition and behavioral development in the preschool child. *Pre-school child malnutrition*. National Health Science, 1966, Public. No. 1282.
Newman, H. H., Freeman, F. N., & Holzinger, K. J. *Twins: A study of heredity and environment*. Chicago: Univ. of Chicago Press, 1937.
Nichols, R. C., & Bilbro, W. C., Jr. The diagnosis of twin zygosity. *Acta genet.*, 1966, 16, 265-275.
Pettigrew, T. *A profile of the Negro American*. Princeton, N. J.: Van Nostrand, 1964.
Powledge, F. *To change a child—A report on the Institute for Developmental Studies*. Chicago: Quadrangle Books, 1967.
Reed, E. W., & Reed, S. C. *Mental retardation: A family study*. Philadelphia: W. B. Saunders Co., 1965.
Research Profile No. 11. Summary of progress in childhood disorders of the brain and nervous system. Washington, D.C.: Public Health Service, 1965.
Reymert, M. L., & Hinton, R. T., Jr. The effect of a change to a relatively superior environment upon the IQs of one hundred children. *Yearb. Nat. Soc. Stud. Educ.*, 1940, 39, (I), 255-268.
Rimland, B. *Infantile autism*. New York: Appleton-Century-Crofts, 1964.
Roberts, J. A. F. The genetics of mental deficiency. *Eugen. Rev.*, 1952, 44, 71-83.
Roberts, R. C. Some concepts and methods in quantitative genetics. In J. Hirsch (Ed.), *Behavior-genetic analysis*. New York: McGraw-Hill, 1967. Pp. 214-257.
Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston, 1968.
Schull, W. J., & Neel, J. V. *The effects of inbreeding on Japanese children*. New York: Harper & Row, 1965.
Schwebel, M. *Who can be educated?* New York: Grove, 1968.
Scott, J. P., & Fuller, J. L. *Genetics and the social behavior of the dog*. Chicago: Univer. of Chicago Press, 1965.
Stodolsky, S. S., & Lesser, G. Learning patterns in the disadvantaged. *Harvard educ. Rev.*, 1967, 37, 546-593.
Scrimshaw, N. S. Infant malnutrition and adult learning. *Saturday Review*, March 16, 1968. p. 64.
Shields, J. *Monozygotic twins brought up apart and brought up together*. London: Oxford Univer. Press, 1962.
Shuey, A. M. *The testing of Negro intelligence*. (2nd ed.) New York: Social Science Press, 1966.
Skeels, H. M. Adult status of children with contrasting early life experiences: A follow-up study. *Child Developm. Monogr.*, 1966, 31, No. 3, Serial No. 105.
Skeels, H. M., & Dye, H. B. A study of the effects of differential stimulation on mentally retarded children. *Proc. Addr. Amer. Ass. Ment. Defic.*, 1939, 44, 114-136.
Spuhler, J. N., & Lindzey, G. Racial differences in behavior. In J. Hirsch (Ed.), *Behavior-genetic analysis*. New York: McGraw-Hill, 1967. Pp. 366-414.
Stoch, M. B., & Smythe, P. M. Does undernutrition during infancy inhibit brain growth and subsequent intellectual development? *Arch. Dis. Childh.*, 1963, 38, 546-552.
Stoddard, G. D. *The meaning of intelligence*. New York: Macmillan, 1943.
Stott, D. H. Interaction of heredity and environment in regard to 'measured intelligence.' *Brit. J. educ. Psychol.*, 1960, 30, 95-102.
Stott, D. H. *Studies of troublesome children*. New York: Humanities Press, 1966.
Thompson, W. R. The inheritance and development of intelligence. *Res. Pub. Ass. Nerv. Ment. Dis.*, 1954, 33, 209-331.
Thorndike, E. L. Measurement of twins. *J. Philos., Psychol., Sci. Meth.*, 1905, 2, 547-553.

Tuddenham, R. D. Psychometricizing Piaget's méthode clinique. Paper read at Amer. Educ. Res. Ass., Chicago, February, 1968.
Tyler, L. E. *The psychology of human differences.* (3rd ed.) New York: Appleton-Century-Crofts, 1965.
U.S. Commission on Civil Rights. *Racial isolation in the public schools.* Vol. 1. Washington, D.C.: U.S. Government Printing Office, 1967.
U.S. News and World Report, Mental tests for 10 million Americans—what they show. October 17, 1966. Pp. 78-80.
Vandenberg, S. G. Contributions of twin research to psychology. *Psychol. Bull.,* 1966, **66**, 327-352.
Vandenberg, S. G. Hereditary factors in psychological variables in man, with a special emphasis on cognition. In J. S. Spuhler (Ed.), *Genetic diversity and human behavior.* Chicago: Aldine, 1967. Pp. 99-133.
Vandenberg, S. G. The nature and nurture of intelligence. In D. C. Glass (Ed.), *Genetics.* New York: The Rockefeller University Press and Russell Sage Foundation, 1968.
Vernon, P. E. Symposium on the effects of coaching and practice in intelligence tests. *Brit. J. educ. Psychol.,* 1954, **24**, 5-8.
Vernon, P. E. Environmental handicaps and intellectual development: Part II and Part III. *Brit. J. educ. Psychol.,* 1965, **35**, 1-22.
Walters, C. E. Comparative development of Negro and white infants. *J. genet. Psychol.,* 1967, **110**, 243-251.
Wheeler, L. R. A comparative study of the intelligence of East Tennessee mountain children. *J. educ. Psychol.,* 1942, **33**, 321-334.
White, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 2. New York: Academic Press, 1965. Pp. 187-220.
Willerman, L., & Churchill, J. A. Intelligence and birth weight in identical twins. *Child Developm.,* 1967, **38**, 623-629.
Wilson, A. B. Educational consequences of segregation in a California community. In *Racial isolation in the public schools,* Appendices, Vol. 2 of a report by the U.S. Commission on Civil Rights. Washington, D.C.: U.S. Government Printing Office, 1967.
Wiseman, S. *Education and environment.* Manchester: Manchester Univer. Press, 1964.
Wiseman, S. Environmental and innate factors and educational attainment. In J. E. Meade & A. S. Parkes (Eds.), *Genetic and environmental factors in human ability.* New York: Plenum Press, 1966. Pp. 64-80.
Wright, S. Statistical methods in biology. *J. Amer. stat. Ass.,* 1931, **26**, 155-163.
Young, M., & Gibson, J. B. Social mobility and fertility. In J. E. Meade & A. S. Parkes (Eds.), *Biological aspects of social problems.* Edinburgh: Oliver and Boyd, 1965.
Zazzo, R. *Les jumeaux, le couple et la personne.* 2 vols., Paris: Presses Universitaires de France, 1960.
Zigler, E. Familial mental retardation: a continuing dilemma. *Science,* 1967, **155**, 292-298.
Zigler, E. The nature-nurture issue reconsidered: A discussion of Uzgiris' paper. Paper read at Conference on Sociocultural Aspects of Mental Retardation, Peabody College, Nashville, Tenn., June, 1968.

A Two-Factor Theory of Familial Mental Retardation

Arthur R. Jensen

University of California, Berkeley

Varieties of Mental Retardation

Persons who are validly classified as mentally retarded are a highly diverse group. Not only is there great variety in the behavioral and emotional characteristics of the retarded, or in their social adaptability, but careful examination reveals marked difference among them even in their mental abilities, both quantitatively and qualitatively. The causes of mental retardation, as well as their behavioral manifestations, are also diverse. It is our task scientifically to understand this diversity.

Two broad categories of mental deficiency are now generally recognized. The first is comprised of those conditions resulting from (a) chromosomal anomalies (e.g., Down's syndrome or "Mongolism"); (b) major gene defects whereby a single mutant gene, usually recessive, completely overrides the normal determinants of mental development (e.g., phenylketonuria and microcephaly), (c) brain damage due to infectious disease or trauma (e.g., maternal rubella, encephalitis, eclampsia). The vast majority of the most severely retarded, with IQs below 50, belong in this category.

The second category consists of what is now called familial mental retardation. The vast majority of these individuals are mildly retarded, with IQs between 50 and 70. (The upper limit seems quite arbitrary and has been placed anywhere from 70 to 85). At least 80 to 90 percent of persons in this IQ range appear clinically normal and show no history or signs of neurological damage.

Paper presented at the 4th International Congress of Human Genetics, Paris, France, September 9, 1971.

The first category of retardation, although it is continuous with the normal distribution of intelligence in the population, is in a sense separate from it. It is superimposed upon the normal distribution and creates the "bulge" at the lower tail of the distribution, that is, the excess over the frequency of low IQs that would be expected from a polygenic and microenvironmental model of the distribution of intelligence in the population.

The second category of retardation, that is, the so-called familial variety, on the other hand, can be viewed as just the lower tail (about 3 percent) of the normal distribution. Although such retardation constitutes normal variation rather than a pathological condition, for the individual it is usually a severe handicap educationally, occupationally, and socially. Such persons as adults in a modern industrial society can seldom manage on their own and they usually require various social services for their welfare.

The relative frequency of this "normal," aclinical kind of retardation increases drastically as we move from higher to lower socioeconomic segments of the population. But it is worth noting that when the mean IQ of all individuals within any given stratum of the population (or a random sample thereof) is determined, this information alone permits a considerably accurate estimate of the frequency of familial retardation within that segment of the population. (The same thing holds true in predicting the frequency of intellectually gifted persons in a given segment of the population.) It is clear that there is a highly lawful relationship between the overall mean IQ of a population (or subgroup thereof) and the frequency of familial retardation (and of giftedness) in that population. This is even more true when we consider only the children in a population group rather than the parental generation, which, with the high degree of social

mobility found in modern Western societies, has already become quite sorted out along occupational and socioeconomic lines, creating markedly skewed IQ distributions of the adult populations of the upper and lower SES groups. These quite lawful relationships to which I have just referred are best described in terms of the properties of the normal curve, and they are highly consistent with a polygenic theory of the distribution of intelligence. Mental retardation of the second kind, therefore, cannot properly be regarded in isolation from parameters of the intelligence distribution in the whole population. Some theories of the etiology of mental retardation and the programs proposed for its amelioration all too often overlook this central fact.

Heterogeneity of Abilities in Familial Retardation

From here on I shall be concerned only with the aclinical variety of mental retardation, which accounts for at least 80 percent of all mental deficiency. Although individuals in this category span a range of 20 or 30 IQ points and differ in predictable ways as a result, there are actually greater ability differences within this group than one might expect on the basis of IQ differences alone. Indeed, retarded persons having the very same IQ are often seen to differ quite markedly in their abilities, and to differ in ways that do not seem entirely accountable in terms of differences in their experience and training. Children and adults in the IQ range from 50 to 80 are known to differ greatly in vocational aptitudes, in social adaptability, and in various non-scholastic aptitudes, and these differences are only slightly related to their IQs obtained on the best standard tests. What these persons share most in common is an inordinate difficulty in regular school work. Under the usual conditions of class instruction, they lag far behind the average child of the same age, and the gap increases from

earlier to later years. In non-academic pursuits, on the other hand, these children show great diversity of ability. I have been concerned with understanding the basis for this diversity. Although my present conclusions have taken shape gradually throughout a series of empirical studies, it will be most efficient to begin by presenting the main points of my formulation as it now stands.

The Basic Observations

There are several inter-related empirical observations which my theoretical formulation attempts to explain.

First, there is the fact that retarded children in the IQ range between 50 and 80 are a relatively homogeneous group in performance on practically all standard intelligence tests. Most individual tests, such as the Stanford-Binet and the Wechsler scales have their highest reliability and concurrent validity in this range of the IQ distribution.

Second, there is the fact that within this rather homogeneous group with respect to IQ, there is apparently a very much greater range of other abilities, including cognitive abilities, provided they are non-academic in the traditional sense of the word. These abilities have been noted in the casual observations of parents, teachers, school psychologists, and the like, as great differences in the acquisition of skills on the playground, in social skills, and in practical knowledge and shrewdness in coping with the environment.

Third, there is the fact that children of the lowest socioeconomic status (SES), who comprise by far the largest proportion of the aclinical mentally retarded, show the greatest discrepancy, on the average, between their low IQs and these other kinds of abilities I have referred to. This seems especially true of Negro children of low SES. Middle-class white

children with low IQs, on the other hand, generally show a more all-round retardation. Their poor performance on IQ tests is more consistent with their general behaviour, in and out of school, than seems to be the case with low SES retarded children, whose mental handicap often seems confined almost entirely to the more academic aspects of schooling.

These casual observations by teachers and school psychologists have contributed largely to the popular belief that the standard IQ tests are somehow culturally biased against children of low SES and in favor of middle-class white children. The tests are seen as seriously underestimating the intelligence of low SES children. The fact that the IQ predicts scholastic performance equally well for low SES as for middle SES children is usually explained away by saying that schooling itself, both the academic curricula and the methods of instruction, is culturally biased in favor of the middle class. Until a few years ago I had subscribed completely to this commonly held viewpoint, and my research in this area actually began with an attempt to formalize these observations in the psychological laboratory and thereby to demonstrate, by more precise and rigorous methods than had yet been applied, that the much higher incidence of retardation among children of low SES, particularly among minority children, was the fault of the IQ tests and also, possibly, of the schools. My own research in this vein has since led me to reject this view. But the theory I have gradually arrived at to replace it is quite different from the simple alternative that existed before I began my research.*

In order to analyze the basic observations which I have just described, a series of laboratory studies were conducted in which we compared retarded and average children of lower and middle SES (including Negro, Mexican, and white children) on a number of standard IQ tests and also on a considerable variety of other cognitive tasks. (We were not interested in sensory

and motor skills or other abilities outside the cognitive domain.) These studies have been summarized elsewhere in more detail than is possible here (Jensen, 1968a,b,c; 1969a,b,c; 1970a,b,c; 1971; Jensen & Rohwer, 1968, 1970). What these studies show, aside from any theoretical interpretation, are essentially the following points:

1. On a variety of tests of rote learning and short-term memory, retarded children score much less far below children of average IQ than on tests involving abstraction, reasoning, problem solving, and conceptual learning. Consequently, some considerable proportion of children who are retarded in terms of IQ are able to perform at an average level or above on a certain class of tasks that clearly involve mental ability. These are represented in our laboratory studies by (a) Trial-and-error selective learning with visual and auditory reinforcements for correct responses. (These problems have involved the trial-and-error acquisition of anywhere from 2 to 12 S-R associations.) (b) Serial rote learning, using lists of familiar objects (e.g., cup, comb, pencil, etc.), pictures of familiar objects, colored geometric forms, nonsense syllables, and common nouns. (c) Paired-associates learning, using the same or similar materials as in the serial learning. (d) Free recall learning (e.g., presenting 20 familiar objects and asking the subject to recall, in any order they come to mind, the names of as many of the items as possible when they are put out of sight), using the same materials as above. (e) Digit span memory under different conditions of presentation and recall (e.g., recall immediately after presentation of the string of digits; recall 10-seconds after presentation; and recall after three successive presentations of the same string of digits.

What all these tasks have in common, as contrasted with tasks on which all retardates perform much more poorly, is that they call for little or no transformation of the stimulus input in order for the subject to arrive at the response output. Stimulus and response are highly similar. What the tasks call for essentially is accurate registration of sensory experiences, immediately giving already well-learned names or labels to these, and at some later point in time repeating these labels in response to partial stimulus cues. It is a kind of recording and playback on cue, as contrasted with the other class of cognitive tasks, those on which retardates perform most poorly, involving transformation and mental manipulation of the input in order to produce the answer -- the relating and comparing of present stimuli with past learning, generalization and transfer of old learning to the new problem, the abstraction of conceptual and semantic similarities and differences, etc. All of these latter processes especially characterize those kinds of intelligence test items which are most highly loaded with *g*, the general factor common to all intelligence tests, which Spearman characterized as an ability for the "eduction of relations and correlates." For convenience I have labelled these two broad types of mental ability Level I (for non-transformational learning and retention) and Level II (for intelligence as characterized by *g*).

2. Level I and Level II abilities show an interaction with SES such that retarded low SES children are on the average superior in Level I ability to middle SES children of the same IQ. Those retardates who appear most adequate in non-academic activities are generally average or above average in Level I. It is not uncommon, for example, to find low SES Negro children with IQs below 60 who perform in the average range or above on Level I tests. Yet their counterparts in this respect are exceedingly rare among low IQ middle and upper-middle class white children, who almost always perform well

below the average on Level I tests.

Institutionalized retardates (and usually those in "sheltered workshops"), as contrasted with a representative sample of all retardates in the population, are usually low both in Level I and Level II abilities. It is therefore doubtful if my findings would ever have been made had I tested only institutionalized individuals. There are marked differences between retardates who become more or less self-sufficient out in the world and those who must be cared for. Psychometrically this difference is not much related to IQ but is more markedly related to Level I ability.

In attempting to understand these findings, our first thought was that the Level II tests were more culturally biased against low SES individuals and that therefore, for any given IQ, the low SES person was really more intelligent than the high SES person, and this difference would show up in the presumably less culture-biased Level I tests. In short, I at first thought I had found in my Level I tests a culture-free or a culture-fair means of measuring intelligence. But this idea has proved to be wrong. A variety of Level II tests differing in degree of culture-loading all show highly consistent results: We have found no tests, verbal or nonverbal, with any appreciable complexity or substantial _g_ loading on which properly diagnosed retarded children score in the average range. And surprisingly enough, low SES children, especially if they are Negro, actually score slightly higher on the verbal and the more obviously culture-loaded tests than on nonverbal tests of the type that attempt to minimize middle-class cultural content. Also, the experimental manipulation of task variables in laboratory learning experiments so as to either minimize or maximize the role of Level II processes leads me to the conclusion that the Level I - Level II distinction is not a matter of the culture-loading of the tests that measure each type of ability but of the different kinds of mental

processes required in the two classes of tests. Nor is the difficulty of the task the essential basis of distinction. Level I and Level II test items can be made equally difficult in terms of their p values (i.e., the percentage of the population that can perform successfully). The essential distinction between Level I and Level II is in the complexity of the mental transformations required for successful performance on the task. Moreover, twin and sibling correlations and estimates of the heritability (i.e., the proportion of the total variance in test scores attributable to genetic factors) of Level I and Level II tests give no indication of significantly lower heritability of Level II than of Level I tests. If Level II tests reflect environmental or cultural influences to a greater extent than Level I tests, one should expect lower heritability values for Level II tests. But this is not the case, and, if anything, slightly the reverse seems to be true.

Level I and Level II in the General Population

In order to determine just how far below the average of the population retarded children stand on Level I tests, we have given such tests to large, representative samples of the school age population, now totalling 15,000 children in all. And to study the relationship between Level I and Level II abilities, verbal and nonverbal intelligence tests, representative of Level II, have also been administered to the same large samples. These large-scale data obtained from the general population put our findings with the mentally retarded into a proper perspective and show that they are not isolated phenomena peculiar to retardates but are a consequence of certain population characteristics.

The regression of Level I test scores on IQ or Level II scores in all samples appears to be linear throughout the IQ range from about 50 to 150.

The slope of the regression line and the correlation between Level I and Level II abilities differs from one subpopulation group to another. It is lower in low SES groups and higher in upper SES groups. It is especially lower among Negroes as compared with whites. In various studies the correlation between Levels I and II have ranged from .10 to .40 in low SES groups, comprised largely of Negro children, and from .50 to .70 in middle SES groups comprised largely of white children. (However, a sample of Oriental-American children, although of lower SES than the white sample, showed an even higher correlation between Levels I and II than was found in the white sample.) Because the regression of Level I on Level II has a steeper slope (higher correlation) in higher than in lower SES groups, the regression lines of lower and upper SES groups must inevitably cross. Consequently, in the region of low IQ that characterizes mental retardation, the lower SES group obtains higher average scores on Level I tests -- which is the phenomenon described earlier. These relationships are shown in Figure 1.

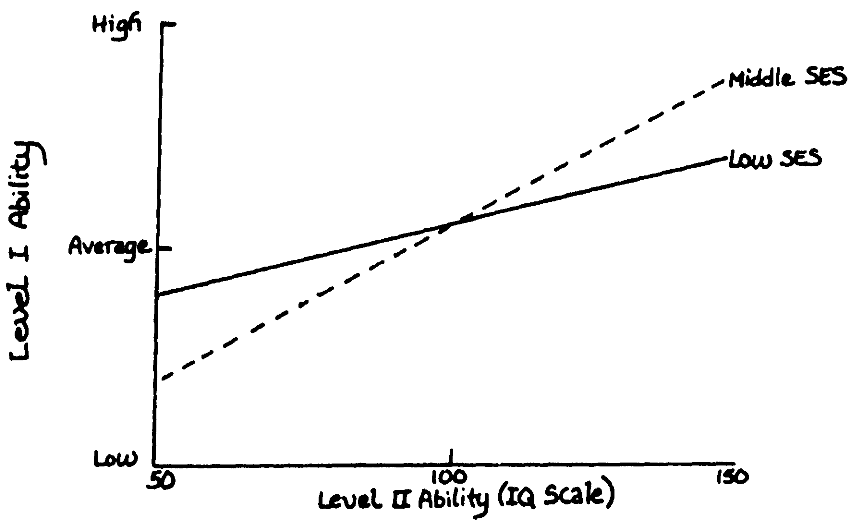


Fig. 1. Typical regression lines of Level I upon Level II ability in middle and low socioeconomic groups.

Thus, the phenomenon of higher Level I ability among lower than among upper SES retardates, on the average, is seen to be a consequence of the lower correlation between Levels I and II in the low SES group as compared with the higher SES group. But what we did not expect to find before we finally tested children in adequately large numbers throughout the entire range of IQ is the reverse phenomenon at the upper end of the IQ scale, that is, the finding that low SES children (most of whom are Negro in these studies) with high IQs perform significantly less well than their middle SES counterparts in IQ. This came as something of a surprise, but it is now based on such substantial evidence that its factual status is beyond reasonable doubt. From a theoretical standpoint it is, of course, a simpler, more lawful picture than we would have if the regression were not linear and the consequent reverse symmetry at the low and high ends of the IQ scale did not obtain.

This finding, furthermore, helps to clarify a point about which there was some doubt in the earlier stages of our research. This was the question of whether low SES retardates performed better on Level I tests, relative to those of middle SES, simply because Level I tests were less culturally biased than the IQ tests. This culture-bias hypothesis seems untenable in view of the fact that in the range of IQ above 100, low SES children perform relatively less well on Level I tests. Also, when we have given various Level II tests which differ obviously in culture loading, such as the Peabody Picture Vocabulary Test and Raven's Progressive Matrices, and then have examined the regression of the less culture-loaded on the more culture-loaded test, we find no cross-over of the regression lines of the low and middle SES groups; the lines are quite parallel. In short, comparison of lower and upper SES groups on Level I vs. Level II

tests gives a quite different picture from that of comparing the two groups on culture-loaded vs. culture-fair tests.

Nature of the Relationship Between Levels I and II

Does the correlation between Level I and Level II abilities represent a functional dependence of Level II upon Level I? For example, is above-average Level I ability a necessary but not sufficient condition for above-average Level II ability in the sense, say, that knowledge of subtraction is a necessary but not sufficient condition for solving problems in long division? Obviously some degree of learning and memory (i.e., Level I ability) are essential for intellectual development. But above some low threshold of Level I ability, is there any functional dependence of individual differences in Level II upon individual differences in Level I? We know, of course, that there is some correlation, often quite substantial, between Levels I and II. But correlation does not necessarily imply functional dependence of one set of processes upon another, in this case Level II upon Level I. This question has puzzled us for some time. It probably cannot be answered definitively on the basis of the evidence now available. A number of lines of evidence, however, suggest a hypothesis that seems most likely to be true.

In the first place, the wide range of correlations between Levels I and II, going from .20 to .80 (after corrections for attenuation and restriction of range) in various subpopulations, seems inconsistent with a high degree of functional dependence between the two types of ability. If the correlation were completely a result of functional dependence, it is difficult to see why the dependency should be so much higher in one population group than in another. Secondly, a high degree of functional dependence

would imply an increasing correlation between Levels I and II with increasing age from early childhood to early maturity, since this is the period of marked development of Level II abilities. But we have found no evidence of greater correlation between Levels I and II with increasing age, and slightly the opposite is the case. Subjects with high IQs but low Level I ability are somewhat less common among younger children between the ages 4 and 7 than among children beyond 10 years of age. It is as if Level I ability acts as scaffolding for the develop[illegible] of Level II abilities and then falls away in importance as the Level II abilities are consolidated. The child who is below average in Level I and above average in Level II will appear to be a slow developer in Level II in early childhood; he is in a sense a slow learner who, because of good Level II ability, is able thoroughly to understand and consolidate everything he learns and incorporate it into the cognitive structures we call intelligence. Later in development these Level II cognitive structures become relatively more important in educational attainments, and the child who is relatively low in Level I but high in Level II becomes much less handicapped in school than the child who shows the opposite pattern of abilities. The low I - high II child is one who learns with difficulty in school when the learning is more or less rote and affords little opportunity to grasp concepts and relationships; he is slow in acquiring skill that requires sheer repetition; but once it has been acquired, he can fully bring to bear what he has learned in logical reasoning and problem solving. He <u>understands</u> what he learns, though he may have learned it slowly. Such children, who often seem to get off to a slow start in the early grades in school, appear to become brighter and intellectually more capable as they progress in school and as the academic subject matter makes increasing demands on conceptual and

abstract thinking and involves relatively less sheer acquisition of simple skills and factual information. The high I - low II child, on the other hand, presents a very different picture. In early childhood he may appear quite bright and quick in picking up all kinds of simple skills and verbal knowledge; he may appear linguistically precocious; he may do quite well in scholastic subjects and skills that depend upon learning by repetition such as penmanship, spelling, mechanical arithmetic, memorizing the words of songs, etc., but he experiences increasing difficulty and frustration -- sometimes to the point of hating school -- as the conceptual and abstract demands of the subject matter increase from earlier to later grades. It becomes increasingly difficult to understand what is learned, and, when ultimately in some academic subjects learning and understanding become one and the same, the pupil with a marked deficiency in Level II is almost totally handicapped. While one can find some small percentage of pupils of below-average Level I ability who are doing very well, say, in algebra or science, there are virtually no below-average Level II pupils who are succeeding in these subjects.

If there is at most only a slight degree of functional dependence of Level II upon Level I, as suggested by the fact that some few older children with high Level II ability are found to be well below-average in Level I, what is the basis for the correlation between Levels I and II and for the fact that it differs so markedly in different populations? The most plausible explanation is in terms of genetic assortment. If Levels I and II are controlled by two different polygenic systems, these can become assorted together to any degree in a given population through selective and assortative mating. I have rejected the idea that only Level I ability is genetically determined and that Level II abilities are learned, acquired, or developed out of Level I abilities entirely as a result of environmental

influences. If this were the case, the heritability of intelligence (Level II) should not be as high as we know it to be -- about 70 to 80 in present-day populations. Also, according to this notion, Level I should have much higher heritability than Level II. But the correlations obtained on siblings and twins give no indication that Level I abilities are significantly more heritable than Level II abilities, and if anything, Level I ability appears less heritable than Level II. It seems much more likely that both Level I and Level II are controlled by distinct polygenic systems and are correlated to varying degrees in different population groups because these groups have differed in the kinds of demands that would cause the genetic factor underlying Levels I and II to become assorted together. We know there is a high degree of assortative mating for intelligence in European and North American Caucasian populations. In fact, in Western society there is probably a higher degree of assortative mating for intelligence than for any other trait.

This should not be too surprising since educational attainments, occupational level, and socioeconomic status, which are the basis for assortative mating, are highly correlated with intelligence. If Level I ability also has some correlation with occupational and socioeconomic status independently of intelligence (Level II), we should expect the genetic factors involved in Levels I and II to become associated through assortative mating. This is consistent with the observation that omnibus-type intelligence tests which involve an admixture of both Level I and Level II (e.g., the Stanford-Binet and Wechsler tests) show a higher correlation with practical criteria such as educational achievement and occupational status than do factorially more pure tests of Level II, such as the Raven Matrices. Populations that

have not long been stratified educationally and occupationally would have had less assortative mating for these abilities, and consequently would show a lower correlation between them, as we find, for example, in the American Negro population as contrasted with the white. Also, Level II ability, being more highly related to the academic and intellectual demands of schooling and higher occupational status is more subject to assortative mating and consequently to genetic stratification in terms of socioeconomic status. Good Level I ability, on the other hand, is more or less equally advantageous in all cultures and walks of life and would therefore become less differentiated than Level II among various population groups.

Physiological Basis of Level I and II Abilities

This is quite speculative, but from what we know about the organization of the nervous system it is an interesting hypothesis that the basic locus of Level I abilities is in the electrochemical processes involved in short-term memory and the neural consolidation of memory traces. The biochemical basis of these processes is evinced, for example, in the fact that learning and memory, which involve neural consolidation, can be altered by pharmacological means. Level II abilities, on the other hand, are hypothesized to depend upon the structural aspects of the brain -- the number of neural elements and the complexity and organization of their potential interconnections.

The evolution of the nervous system, represented in the hierarchy of phyla, is most evident in the development of Level II processes. The growth of mental ability in the individual similarly reflects largely the

gradual emergence of Level II processes from infancy to maturity (Jensen, in press). G. Stanley Hall's famous dictum that "ontogeny recapitulates phylogeny" appears to hold true for mental as well as physical development. The growth curves of Level I and II are quite different, with Level I approaching its developmental asymptote at an earlier age than Level II.

Theoretical Overview

The picture is that of a fundamental division of mental abilities into Level I (learning and memory) and Level II (intelligence, i.e., analytic understanding, reasoning, abstraction, conceptual thinking). Individual differences in both Levels I and II are viewed as due mainly to independent polygenic factors. The distributions of Level I and II abilities in the population are approximately normal. The correlation between Levels I and II is due mainly to the common assortment of the genes involved in the two types of ability. (But there is also some moderate degree of functional dependence of Level II upon Level I.) The genetic correlation differs in various subpopulations, being lower in the low SES segment of the population and higher in the middle and upper-middle class segment. The correlation is lower in the American Negro than in the white population. Because education makes greater demands on Level II than on Level I and the occupational hierarchy and socioeconomic status are highly related to educational attainments in Western societies, there is a much greater mean difference between social classes in Level II than Level I. While Level I is distributed about very similar means in lower and upper SES groups, the means of the Level II distributions may differ by one standard deviation or more. (One standard deviation is equivalent to about 15 IQ points.)

Mental retardation of the type which is a part of the normal distribution of abilities in the population can be described as primary retardation if it

involves marked deficiency in both Levels I and II and as secondary retardation if there is a deficiency only in Level II ability. Secondary retardates often appear normally bright and capable of learning and achievement in many situations, although they invariably experience great difficulties in school work under the traditional curricula and methods of instruction. Many secondary retardates who are regarded as backward children while in school later become socially and economically adequate persons once they are out of the academic situation. Primary retardates, on the other hand, appear to be much more handicapped in the world of work. A serious shortcoming of ordinary IQ tests is that they measure predominantly Level II and fail to distinguish between primary and secondary retardation. Tests that reliably measure both Levels I and II should be developed for use in schools, in personnel selection, and in the armed forces. This formulation also has important implications for the education of children now popularly called culturally disadvantaged, most of whom have normal Level I ability but are often quite far below average in Level II. Such children might benefit educationally from instructional methods which make the acquisition of scholastic skills less dependent upon Level II abilities and more fully engage Level I abilities as a means of raising their educational attainments.

References to Jensen's Level I - Level II

Theory of Mental Abilities

Jensen, A. R. Social class, race, and genetics: Implications for education. American Educational Research Journal, 1968, 5, 1-42. (Reprinted in H. F. Clarizio, R. C. Craig, & W. A. Mehrens (Eds.) Contemporary issues in educational psychology. New York: Allyn & Bacon, in press. (Reprinted in: I. J. Gordon (Ed.) Readings in research in developmental psychology. Glenview, Illinois: Scott, Foresman & Co. Pp. 54-67.) a

Jensen, A. R. The culturally disadvantaged and the heredity-environment uncertainty. In J. Hellmuth (Ed.) The culturally disadvantaged child. Vol. 2. Seattle, Wash.: Special Child Publications, 1968. Pp. 29-76. b

Jensen, A. R. Patterns of mental ability and socioeconomic status. Proc. Nat. Acad. Sci., 1968, 60, 1330-1337. c

Jensen, A. R. How much can we boost I.Q. and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123. (Reprinted in Environment, heredity, and intelligence. Harvard Educational Review, Reprint Series No. 2, 1969, pp. 1-123.) a

Jensen, A. R. Intelligence, learning ability, and socioeconomic status. Journal of Special Education, 1969, 3, 23-35. (Reprinted in Mental Health Digest, 1969, 1, 9-12.) b

Jensen, A. R. Jensen's theory of intelligence: A reply. Journal of Educational Psychology, 1969, 60, 427-431. c

Jensen, A. R. Learning ability, intelligence, and educability. In V. Allen (Ed.), Psychological factors in poverty. Chicago: Markham, 1970. Pp. 106-132. a

Jensen, A. R. A theory of primary and secondary familial mental retardation. In N. R. Ellis (Ed.) International Review of Research in Mental Retardation. Vol. 4. New York: Academic Press, 1970. Pp. 33-105. b

Jensen, A. R. Hierarchical theories of mental ability. In B. Dockrell (Ed.), On intelligence. Toronto: Ontario Institute for Studies in Education, 1970, Pp. 119-190. c

Jensen, A. R. Do Schools Cheat Minority Children? Educational Research, 1971.

Jensen, A. R. The phylogeny and ontogeny of intelligence. Perspectives in Biology and Medicine, in press.

Jensen, A. R., & Rohwer, W. D., Jr. Mental retardation, mental age, and learning rate. Journal of Educational Psychology, 1968, 59, 402-403.

Jensen, A. R., & Rohwer, W. D., Jr. An experimental analysis of learning abilities in culturally disadvantaged children. Final Report. Office of Economic Opportunity, Contract No. OEO 2404, 1970. Pp. 1-181.

APRIL 1970/VOLUME XXXIII/NUMBER 6

The Heritability of Intelligence

ARTHUR R. JENSEN

In accounting for the causes of the differences among persons in IQ, the genes outweigh the effects of environment by 2 to 1.

Since the dawn of history people have noticed differences in intelligence among individuals and have wondered about the causes of these obvious differences. Intelligence has been described by many different words—brightness, cleverness, reasoning power, judgment, and quickness in learning, in grasping abstract concepts, and in solving problems. Every parent, teacher, and employer has observed differences among children and adults in all these characteristics that we call "intelligence." A few persons appear extremely "bright," a few appear extremely "dull," and the vast majority falls somewhere between these extremes. There is a continuous gradation of mental ability from the one extreme to the other, from idiot to genius. Just as we see a continuous gradation of differences in other characteristics of humans, such as physical stature, so too there is a similar gradation of differences in intellectual ability. Indeed, individual variation is a fundamental aspect of all living things. Without individual variation, biological evolution as we know it could not have occurred.

The question of why people differ in intelligence has been asked for centuries, but a scientifically acceptable answer did not become wholly possible until psychologists devised techniques for measuring intelligence quantitatively and objectively. The first really useful intelligence test was devised in 1905 by the French psychologist Alfred Binet. Binet's early test was later revised and improved by Lewis Terman at Stanford University; the now-famous test that resulted from these efforts is known as the Stanford-Binet Intelligence Scale. It is still the most widely used test of general intelligence.

There are also many other intelligence tests, and although many of them appear to be quite different from one another, all actually measure much the same general ability. That is to say, if we administer several seemingly quite different intelligence tests to a large number of persons, their scores on all the tests will be in pretty much the same rank order. Those who score high on one test will tend to score high on the others, and those who score low on one test will usually score low on all the others. This fact of correlation among all tests of intelligence led Charles Spearman, the famous English psychologist, to conclude that there is a general factor, "*g*," which is common to all tests of intelligence. We know that it is practically impossible to make up a mental test having any degree of complexity which does not involve "*g*." We can perhaps most clearly characterize "*g*" as an ability for abstract reasoning and problem solving, for seeing relationships, and for grasping concepts.

A person's score on an intelligence test is usually expressed as an IQ (for Intelligence Quotient). The test is standardized in the general population in such a way that the average IQ at any age is set at 100, and the middle 50 percent of the population falls within the so-called average range of IQ's going from 90 to 110.

Significance of the IQ

Can the IQ tell us anything of practical importance? Is it related to our commonsense notions about mental ability as we ordinarily think of it in connection with educational and occupational performance? Yes, indeed, and there is no doubt about it. The massive evidence from psychological, educational, and industrial research, and research in the armed forces, is unequivocal. We know, for example, that no other single fact that we are now able to ascertain about a child gives us a better prediction of his future scholastic performance than his IQ obtained after age 5 or 6. (Below this age IQ tests become less accurate indicators of the child's later mental development, and below 2 or 3 years of age test scores have practically no predictive value.)

The IQ obtained after 9 or 10 years of age also predicts final adult occupational status to almost as high a degree as it predicts scholastic performance. When various occupations are ranked for average income and for the general public's average judgment of the occupation's prestige and desirability, this rank order is found to be highly related to the average IQ level of the persons in these occupations. There is of course a wide spread of IQ's in nearly every occupation, but the *average* IQ of persons within a particular occupation is closely related to that occupation's standing in terms of its average income and the amount of prestige accorded to it by the general public.

One of the most convincing demonstrations that IQ is

related to "real life" indicators of ability was provided in a classic study by Terman and his associates at Stanford University. In the 1920's they selected a total of 1,528 children with Stanford-Binet IQ's above 140. The average IQ of the group was 152. These children were investigated periodically over the years up into their adulthood. (Most of them are now in their 50's.) Terman found that for the most part these high-IQ children in later adulthood markedly excelled the general population on every indicator of achievement that was examined: a higher level of education completed; more scholastic honors and awards, higher occupational status; higher income; production of more articles, books, patents, and other signs of creativity; more entries in *Who's Who*, a lower mortality rate, better physical and mental health; and a lower divorce rate. Also, they have much brighter children than the average, their average IQ is 133, a level which is exceeded by only 2 percent of children in the general population.

Findings such as these establish beyond a doubt that IQ tests measure characteristics that are obviously of considerable importance in our present technological society. To say that the kind of ability measured by intelligence tests is irrelevant or unimportant would be tantamount to repudiating civilization as we know it.

The Causes of IQ Differences

The layman usually asks: "Is intelligence due to heredity *or* environment?" The scientist promptly answers: "Both." Without heredity *and* environment there simply is no intelligence. Obviously every person must have had a biological inheritance of genes from his parents and must have grown in an environment, or he wouldn't even be here to take an IQ test. So, of course, both heredity and environment are essential for the existence of the individual or any of his physical and mental characteristics.

But when scientists actually study this problem, we find that they do not even ask the layman's question. The question to which scientists have sought an answer can be stated as follows: How much of the variation among persons in a given population is attributable to differences in their environments and how much to differences in their genetic endowments?

Numerous studies conducted by psychologists and geneticists over the last 40 or 50 years provide an answer to this question. The answer is unambiguous and is generally agreed upon by all scientists who have considered all the evidence. This evidence strongly supports the conclusion that genetic factors are much more important than environmental influences in accounting for *individual differences* in IQ. How much more important? The evidence indicates that genetic factors account for at least *twice* as much of the variation in IQ's as environmental factors. This conclusion has one main limitation. Since all of the major studies in this field were conducted with samples of Caucasian European and North American populations, we cannot confidently generalize their conclusions to other populations, especially those with very dissimilar environments.

What are the kinds of evidence that lead to the conclusion that genetic differences outweigh environmental differences in accounting for individual differences in IQ? Most of this evidence, as it is found in the scientific literature, depends upon quite technical methods of analysis developed in a specialty known as quantitative genetics or population genetics. Some of these methods were devised originally to analyze the roles of heredity and environment in agriculture and animal breeding.

Experiments in Animal Breeding

Experiments in which we explicitly try to breed for some specific trait give us the most certain evidence that variation in the trait has a genetic component. Psychologists have bred rats for speed of learning mazes, which is a good indicator of rat intelligence. By always mating the fast-learning males with fast-learning females, and mating slow-learning males with slow-learning females, it is possible, within 6 to 10 generations, to produce two quite distinct strains of rats in respect to maze-learning ability. The slowest learning rat of the "bright" strain will learn mazes faster than the fastest rat of the "dull" strain. The two strains will differ markedly in the number of trials they need to learn how to run through a maze efficiently, avoiding the blind alleys. These experiments definitely prove that not only physical characteristics but some behavioral traits as well are largely inherited through the parental genes. Thus we should not be surprised to find in humans that differences in some behavioral characteristics, including intelligence, are a product of genetic inheritance.

Identical Twins Reared Apart

One of the most important lines of evidence for the inheritance of intelligence in humans comes from studies of identical twins who were separated shortly after birth and reared in different homes. Identical twins originate from a single fertilized ovum which splits in the course of

Identical twins reared apart differ, on the average, by only 6 to 7 IQ points.

early development to form two individuals. Each member of the pair of twins therefore has exactly the same complement of genes. Consequently, any difference between the twins must be due entirely to nongenetic or environmental differences.

Twins separated shortly after birth are often reared in families that differ markedly in social class, and the range of environmental differences observed in their foster homes is fairly typical of the environmental variations seen in the general population.

Four major studies of identical twins reared apart, conducted in England, Denmark, and the United States, and totaling 122 pairs of twins, are in remarkably close agreement in showing that twins reared in different homes are still much more alike in IQ than are *fraternal* twins reared together. Fraternal twins are merely siblings who happen to be conceived and born at the same time, and therefore half of them are of opposite sex. In IQ and other traits they resemble one another no more than do ordinary siblings born at different times.

Identical twins reared apart differ, on the average, by only 6 to 7 IQ points. But even if we test the very same person on two occasions a week apart, we find that his test score will vary, on the average, by 2 or 3 IQ points. This is the test's "measurement error." When we eliminate this error from the twin data, we find that the twins differ only 4 or 5 points in IQ. Identical twins reared *together* differ by only 2 or 3 points, not including measurement error. The largest IQ difference ever found in a pair of identical twins reared apart is 24 points. More than 17 percent of siblings reared together differ by more than 24 IQ points. The same is true of fraternal twins. But siblings (and fraternal twins) have only half of their genes in common, and they differ on the average by 12 IQ points (excluding measurement error), even when reared together.

The studies of identical twins show clearly that individuals who are genetically identical are almost as much alike in mental ability as they are alike in physical traits, and this is true even when they have grown up in different environments.

Unrelated Children Reared Together

The opposite situation to identical twins reared apart is that of genetically unrelated children adopted at birth by foster parents and reared together. Such children differ from one another, on the average, by 15 to 16 IQ points (excluding measurement error). Compare this with the 17 to 18 IQ points difference between unrelated children reared in *different* homes, or the 15 to 16 points difference between unrelated children brought up in different homes but in the same socioeconomic class. We see that unrelated children brought up together in the *same* home differ from one another in IQ at least 3 or 4 times more than genetically identical twins reared in *different* homes. And the unrelated children reared together differ almost as much in IQ as unrelated children simply picked at random from different homes.

The IQ's of adopted children also show little or no relationship to the IQ's of their adopting parents, but they are almost as closely related to the IQ's of their natural parents as we find in the case of children who are reared by their natural parents.

Children reared in the common environment of an orphanage differ from one another in IQ to approximately the same degree as children picked at random from the total population. The IQ's of orphanage children who have never known their own parents show almost the same degree of correlation with their parents' level of ability as we find in the case of children reared by their own parents.

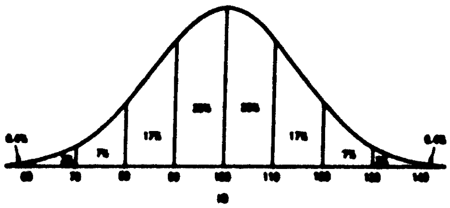
Resemblance Between Parents and Children

Now and then we notice that very bright parents can have an intellectually mediocre child, or that rather dull parents can have an exceptionally bright child. These observations are often pointed to mistakenly as evidence that intelligence is not inherited. But the fact is that genetic theory predicts precisely that we should find such discrepancies between parents and their offspring. For example, parent-offspring differences in height are of about the same relative magnitude as their differences in IQ. Children resemble their parents physically and in mental ability to about the same degree that they resemble their own siblings. The average IQ difference between a parent and his (or her) child is the same as the difference between siblings—that is, about 12 IQ points. The difference between a child and the average of both of his parents' IQ's is about 10 points.

A parent with a high IQ will usually, but by no means always, have children whose IQ's are somewhat lower than his own but are still above the average for the general population. A parent with a low IQ, on the other hand, will usually, but not always, have children whose IQ's are somewhat higher than his own but are still below the average of the population. This phenomenon, discovered by Sir Francis Galton, is called "regression toward the mean," and it holds true for height and other inherited physical traits as well as for IQ.

IQ's of Husbands and Wives

It is interesting that in our society husbands and wives are at least as much alike in IQ as brothers and sisters.



The theoretical normal or Gaussian distribution of IQ's shows the expected percentages of the population in each IQ range. Except at the extremes (below 70 and above 130), these percentages are very close to actual population values. (The percentage figures total slightly more than 100 because of rounding.)

If men and women picked their mates strictly at random, as by a lottery, spouses would differ by an average of 18 IQ points. But in fact men and women choose one another partly for intelligence, and so spouses differ by only 10 or 11 points in IQ.

The Effect of Inbreeding on IQ

Every person harbors a number of mutant, recessive genes. Most of these are defective genes. They are passed on from parent to child, but they usually will not produce any harmful effects to the child unless the other parent also contributes exactly the same defective gene. The reason this usually does not occur is that each parent's normal genes are dominant over the other parent's defective, recessive genes. When mating occurs between a man and a woman who are blood relations, however, the chances are much greater that they will both possess many of the same defective genes. When these defective genes are paired together in the related couple's children, they subtract unfavorably from the traits that are controlled by these genes under normal conditions. This depression due to inbreeding is known to occur in inherited physical traits, such as stature, and the same thing has been found for IQ. It is well established, for example, that cousin marriages produce children who, on the average, have IQ's several points lower than the IQ's of children whose parents are unrelated but are matched with the married cousins on IQ, age, educational level, and socioeconomic status. More extreme are the cases of children who have resulted from incestuous relationships, such as father-daughter and brother-sister matings. These children show a much higher incidence of severe mental retardation than children born to the same parents when they have mated with unrelated persons. These interesting findings are entirely predictable from basic principles of genetics that apply to all living beings. Moreover, it is virtually impossible to explain such facts without concluding that IQ differences are very strongly influenced by genetic mechanisms.

The Relative Effects of Heredity and Environment

How can we summarize briefly what is now known about the relative importance of heredity and environment in causing individual differences in IQ? In the terminology of genetics a summary answer consists of saying that the "heritability" of IQ is close to 0.80. This means that 80 percent of the "variance" in IQ's in the general population is attributable to genetic differences and 20 percent is attributable to nongenetic or environmental differences.

"Variance" is essentially a quantitative index of the total amount of differences that exist among all members of some population. So instead of talking about variance, we can more easily describe our conclusions in terms of average differences.

If we should determine the differences in IQ between every person in the population and every other person, the average of all these differences would turn out to be 18 IQ points. These differences are due both to genetic and to environmental factors. Now we can ask theoretically: What would be the average IQ difference among all persons in the population if everyone had grown up in identical environments from the moment of conception, while genetic differences remained as they are? Under this hypothetical condition of completely equal environments for everyone, the average IQ difference would be 16 points. Thus, there would be a reduction of 2 points in the average difference that now exists. Let us now ask the reverse: What would be the average difference if everyone had exactly the same genetic endowment, but environmental differences remained unchanged? Under this hypothetical condition of complete genetic equality the average IQ difference among persons would be only 8 points, or just half the difference that would exist with equal environments.

So the conclusion we come to—which is certainly valid at least in the white European and North American populations in which the research was conducted—is this: In accounting for the causes of the differences among persons in IQ, the genes outweigh the effects of environment by 2 to 1. As environmental conditions are improved and made more alike for all persons in the society, the average intelligence level of the population will be somewhat increased, and the IQ differences among persons will be slightly reduced. But of course the differences that remain will inevitably be due even more to genetic factors.

TECHNICAL COMMENTS ON SCARR-SALAPATEK'S "RACE, SOCIAL CLASS, AND IQ"

By Arthur R. Jensen

Sandra Scarr-Salapatek's article on "Race, Social Class, and IQ" (1) is to be commended for addressing an important problem in human behavior-genetics, being the first large-scale study of the mental abilities of twins in an American Negro population. Scarr-Salapatek's data, however, raise crucial problems of interpretation which are not even hinted at in her article.

Scarr-Salapatek emphasizes the point that the heritability of the mental tests is in general lower in her Negro sample than in the white. This fact she apparently interprets as being consistent with an explanation of the mean Negro-white IQ differences in terms of environmental factors such as cultural deprivation. She states: "The lower mean scores of disadvantaged children of both races can be explained in large part by the lower genetic variance in their scores" (p. 1293). She adds: "If most black children have limited experience with environmental features relevant to the development of scholastic skills, then genetic variation will not be as prominent a source of individual phenotypic variation; nor will other between-family differences such as SES [socioeconomic] level be as important as they are in a white population" (p. 1294).

The data shown in Scarr-Salapatek's Table 3 (p. 1288), however, make this interpretation highly questionable. These data allow comparison of the mean scores on the combined aptitude tests for Negro children whose parents' level of education and income are both <u>above</u> the median (of the Negro and white samples combined) with the mean scores of white children whose parents' education and income are both <u>below</u> the common median. The

lower status white children still score higher than the upper status Negro children on both the verbal and the non-verbal tests. Although non-verbal tests are generally considered to be less culture-biased than verbal tests, it is the non-verbal tests which in fact show the greater discrepancy in this comparison, with the lower status whites scoring higher than the upper status Negroes. But in this comparison it is the upper status Negro group that has the higher heritability (i.e., greater genetic variance) on both the verbal and non-verbal tests. Thus the lower heritability which Scarr-Salapatek invokes to infer that Negroes' generally poorer performance is attributable to environmental deprivation applies to the lower status white group in this particular comparison. Yet the lower status white group out-performs the upper status Negro group, which has the highest heritability of any of the subgroups in this study (see Table 9, p. 1292).

Is this finding more difficult to reconcile with a strictly environmental explanation of the mean racial difference in test scores than with a genetic interpretation which invokes the well-established phenomenon of regression toward the population mean? In another recent article in Science, Scarr-Salapatek (2) clearly explicated this relevant genetic prediction, as follows:

> "Regression effects can be predicted to differ for blacks and whites if the two races indeed have genetically different population means. If the population mean for blacks is 15 IQ points lower than that of whites, then the offspring of high-IQ black parents should show greater regression (toward a lower population mean) than the offspring of whites of equally high IQ. Similarly, the offspring of low-IQ black parents should show less regression than those of white parents of equally low IQ."

In other words, on the average, an offspring genetically is closer to its population mean than are its parents, and by a fairly precise amount (3). Accordingly, it would be predicted that upper status Negro children should, on the average; regress downward toward the Negro population mean IQ of about 85, while lower status white children would regress upward toward the white population mean of about 100. In the downward and upward regression, the two groups' means could cross each other, the lower status whites thereby being slightly above the upper status Negroes. Scarr-Salapatek's data (Table 3) are quite consistent with this prediction. Scarr-Salapatek's finding is not a fluke; the same phenomenon has been found in other large-scale studies reviewed elsewhere (4).

The exceedingly low (and often even zero or negative) heritability values in the lower SES groups (both Negro and white) raise the question of the adequacy of the particular tests and testing procedures and their comparability with those in other major studies of the heritability of intelligence (5). It is a reasonable conjecture that these group-administered tests, designed for particular grade levels rather than for the full range of ability, were too difficult to allow much spread of scores for many of the low SES children and are therefore less reliable for the low SES than for the higher SES groups. Consistent with this conjecture is the fact that the intercorrelations among the several tests are lower for subjects below the SES median than for those above the SES median (see Table 4, p.1289). And what is the effect of truncating the range by excluding 99 twin pairs who were in retarded classes, fourth-fifths of whom were Negro? (1, Footnote 26). Could this differentially attenuate correlations for Negro and white twins so as to artifactually bolster Scarr-Salapatek's main conclusions? Moreover, we do not know how much the intercorrelations are

affected by the fact that two or more of the tests were administered by the same teacher. I have found it necessary in my own research not to use teacher administered tests, especially in correlational analyses and heritability studies. There are several reasons for this: when children are tested in intact classroom groups, any biasing influence on test performance, such as poorly given instructions, lax observance of the time limits for standardized tests, etc., contributes markedly to "between-teacher" variance, which is tantamount to error variance. It also contributes to spurious intraclass correlation among children in the same class and to spurious inflation of the intercorrelations among subtests, such as the verbal and non-verbal parts. For this reason I have deemed it necessary to employ specially trained testers to insure highly standardized administration of tests, and no more than one test is given by a particular tester, thereby ruling out the possibility of spurious intercorrelation due to any consistent tester bias. The use of teacher administered tests gathered from school files can serve to suggest hypotheses, as in Scarr-Salapatek's study, and are useful in the planning of more thorough studies; but I doubt that they can serve adequately for testing the kinds of quantitative hypotheses put forth by Scarr-Salapatek.

These hypotheses are indeed important and deserve the most thorough investigation possible. Does the complete lack of any statistical tests of significance for the hypotheses considered by Scarr-Salapatek mean that she did not really intend this as an hypothesis-testing study, which many readers are apt to mistake it for? Her approach might be called "weak inference" in contrast to the "strong inference" recommended by Platt (6) and of which investigation in this sphere is so greatly in need.

Institute of Human Learning
University of California
Berkeley, California 94720

Arthur R. Jensen

References and Notes

1. S. Scarr-Salapatek, Science, 174, 1285 (1971).
2. S. Scarr-Salapatek, Science, 174, 1223, p. 1226.
3. The genetic model predicts a parent-child regression of about .50 (or slightly higher if there is assortative mating), and the midparent-child regression is $.50^{\frac{1}{2}}$.
4. A. R. Jensen, Harvard Educational Review, 39, 1 (1969), pp. 83-84.
5. L. Erlenmeyer-Kimling and L. F. Jarvik, Science, 142, 1477 (1963); A. R. Jensen, Proc. Nat. Acad. Sci. U.S., 58, 149 (1967); A. R. Jensen, Behavior Genetics, 1, 133 (1970).

Can We and Should We Study Race Differences?

ARTHUR R. JENSEN
INSTITUTE OF HUMAN LEARNING
UNIVERSITY OF CALIFORNIA, BERKELEY

MOST PERSONS experience some difficulty in discussing the topic of race differences in intelligence—a difficulty over and above that which is ordinarily inherent in the scientific study of any complex phenomenon. There is an understandable reluctance to come to grips with the problem, to come to grips with it, that is to say, in the same straightforward way that we would try to approach the investigation of any other problems in the behavioral sciences. This reluctance is manifested in a variety of "symptoms" found in most writings and discussions of the psychology of race differences, particularly differences in mental ability: a tendency to remain on the remotest fringes of the subject, to sidestep central questions; to blur the issues and tolerate a degree of vagueness in definitions, concepts, and inferences that would be unseemly in any other realm of scientific discourse; to express an unwarranted degree of skepticism about reasonably well established quantitative methods and measurements; to deny or belittle already generally accepted facts—accepted, that is, when brought to bear on inferences outside the realm of race differences; to demand practically impossible criteria of certainty before even seriously proposing or investigating genetic hypotheses, as contrasted with extremely uncritical attitudes toward purely environmental hypotheses, a failure to distinguish clearly between scientifically answerable aspects of the question and the moral, political, and social policy issues; a tendency to beat dead horses and to set up straw men on what is represented as the genetic side of the argument; appeals to the notion that the topic is either really too unimportant to be worthy of scientific curiosity or too complex, or too difficult, or even forever impossible for any kind of research to be feasible, or that answers to key questions are fundamentally "unknowable" in any scientifically acceptable sense; and, finally, the complete denial of intelligence and race as realities, or as quantifiable attributes, or as variables capable of being related to one another, thereby dismissing the subject altogether.

These tendencies will be increasingly overcome the more widely and openly the subject is discussed among scientists and scholars. As some of the taboos against the public discussion of the topic fall away, the issues will become clarified on a rational basis, we will come to know better just what we do and do not yet know about the subject, and we will be in a better position to deal with it objectively and constructively.

IS INTELLIGENCE AN ATTRIBUTE?

Intelligence *is* an attribute of persons. Probably for as long as man has been on earth it has been a common observation that persons differ in brightness, in speed of learning, in ability to solve problems, and so on. Parents, teachers, and employers are able roughly to rank children and adults in terms of a subjective impression of brightness or capability, and there is a fairly high agreement among different observers in the rank order they assign in the same groups of children. It is helpful to think of the subjective perception of intelligence as analogous to the subjective perception of temperatures, which is also an attribute. Before the invention of the thermometer, temperature

was a matter of subjective judgment. The invention of the thermometer made it possible to objectify the attribute of temperature, to quantify it, and to measure it with a high degree of reliability. With some important qualifications, the situation is similar in the case of intelligence tests. The most essential difference is that intelligence, unlike temperature, is multidimensional rather than unidimensional. That is to say, there are different varieties of intelligence, so that persons do not maintain the same rank order of ability in every situation or test that we may regard as indicative of intelligence. It so happens that from among the total spectrum of human behaviors that can be regarded as indicative of some kind of "mental ability" in the broadest sense, we have focused on one part of this spectrum in our psychological concept of intelligence. We have emphasized the abilities characterized as conceptual learning, abstract or symbolic reasoning, and abstract or verbal problem solving. These abilities were most emphasized in the composition of intelligence tests because these were the abilities most relevant to the traditional school curriculum and the first practical intelligence tests were devised to predict scholastic performance. When tests were devised to predict occupational performance, they naturally had a good deal in common with the tests devised for scholastic prediction, since the educational system is intimately related to the occupational demands of a given society. Much the same abilities and skills that are important in schooling, therefore, are important also occupationally. Thus, we find that in industrialized countries practically all intelligence tests, scholastic aptitude tests, military classification tests, vocational aptitude tests, and the like, are quite similar in composition and that the scores obtained on them are all quite substantially intercorrelated. In short, there is a large general factor, or *g*, which the tests share in common and which principally accounts for the variance among individuals. When tests are devised to measure this *g* factor as purely as possible (i.e., in a factor analysis including a host of other tests it will have nearly all of its variance loaded on the general factor common to all the other tests and have little or no variance loaded on factors found only in certain tests [special factors] or factors found only in small groups of tests [group factors]), examination of their item content leads to the characterization of it as requiring an ability for abstract reasoning and problem solving. Raven's Progressive Matrices Test is an example of such a test. Tests having quite diverse forms can have equally high loadings on the *g* factor—for example, the verbal similarities and block design tests of the Wechsler Intelligence Scales are both highly loaded on *g*. Tests of *g* can be relatively high or relatively low in degree of "culture fairness." (The question "In what way are a wheel and a penny alike?" is probably more culture fair than the question "In what way are an oboe and a bassoon alike?") In short, it is possible to assess essentially the same intelligence by a great variety of means.

Standard IQ tests measure the kinds of behavior in abstract and verbal problem situations that we call abstract reasoning ability. These tests measure more of *g*—the factor common to various forms of intelligence tests—than of any of the other more special ability factors, such as verbal fluency, spatial-perceptual ability, sensory abilities, or mechanical, musical, or artistic abilities, or what might be called social judgment or sensitivity. But a test that measured everything at once would not be very useful. IQ tests do reliably measure one important, though limited, aspect of human performance. The IQ qualifies as an appropriate datum for scientific study. If we are to study intelligence, we are ahead if we can measure it. Our measure is the IQ, obtained on tests which meet certain standards, one of which is a high *g* loading when factor analyzed among other tests. To object to this procedure by arguing that the IQ cannot be regarded as being interchangeable with intelligence, or that intelligence cannot really be measured, or that IQ is not the same as intelligence, is to get bogged down in a semantic morass. It is equivalent to arguing that a column of mercury in a glass tube cannot be regarded as synonymous with temperature, or that temperature cannot really be measured with a thermometer. If the measurements are reliable and reproducible, and the operations by which they are obtained can be objectively agreed upon, this is all that need be required for them to qualify as proper scientific data. We know that individually administered IQ tests have quite high reliability; the reliability coefficients are around .95, which means that only about 5% of the total individual differences variance is attributable to measurement error. And standard group administered tests have reliabilities close to .90.

The standard error of measurement (which is about ±5 for the Stanford-Binet and similar tests) must always be taken into consideration when considering any individual's score on a test. But it is

actually quite unimportant in comparison of the means of large groups of subjects, since errors of measurement are more or less normally distributed about zero and they cancel out when N is large. The reliability (i.e., consistency or freedom from errors of measurement) per se of the IQ is really not seriously at issue in making comparisons between racial groups. If the samples are large, the mean difference between groups will not include the test's errors of measurement.

The validity or importance of the measures derives entirely from their relationship to other variables and the importance we attach to them.

The IQ correlates with many external criteria, and at the most general level it may be regarded as a measure of the ability to compete in our society in ways that have economic and social consequences for the individual. In the first place, the IQ accords w h parents' and teachers' subjective assessments of children's brightness, as well as with the evaluations of children's own peers. In terms of assessments of scholastic performance, whether measured in terms of school grades, teachers' ratings, or objective tests of scholastic achievement, the IQ accounts for more of the individual differences variance than any other single measurable attribute of the child. IQ accounts for about 50% of the variance in scholastic achievement at any single grade level, and over the course of several years or more of schooling it accounts for over 70% of the variance in overall scholastic performance.

Since considerably less than 100 percent of the variance is accounted for, it means the IQ is not an infallible predictor of the performance of any one individual. When used for individual diagnosis it must be evaluated in terms of many other factors in the child's makeup and background and condition at the time of testing, and even then not too much stock should be placed in the IQ in predicting *for the individual case*, since the predictive validity of the IQ is not sufficiently high to override the effects of possibly unassessed traits or unpredictable unusual future circumstances which may radically alter the course of the individual's development or performance in a statistically small proportion of cases. Thus, I am emphasizing the importance of evaluating the IQ somewhat differently when used for individual diagnosis and prediction than when used in making statistical predictions on large groups of individuals. It is somewhat analogous to actuarial predictions of insurance risks. Predictions for large groups classified by various criteria can be made with high degrees of certainty, while predictions for individual cases are highly uncertain.

Recently I received a letter from a high school senior who described himself as coming from a disadvantaged background. He had a strong desire to go on to college in hopes of becoming a lawyer, and he was wondering about his IQ and how much stock he should put in it in deciding his further course. I doubt if there is much more sense in worrying about one's own IQ than in worrying about the age at which one will die, as predicted by the insurance company's actuarial tables. Among other things, I wrote the following to my student inquirer: "My own attitude toward tests, when I was a student, was not to give much thought to them but simply to set my sights on what seemed to me a realistic goal and then do my best to achieve it. You find out from those who have already made it what you have to know, what you have to be able to do, what skills you need to develop, and you set about doing these things just as you'd go about doing any kind of job that you know has to be done. If you set your goals too low, it's too easy and you won't develop your potential. If you set goals that are unrealistically high, you become discouraged. I recommend one step at a time, each step being something you really think you can achieve if you really work for it. When you have made the first step successfully, then you will have a better idea of how to take the next step. That way, if you have whatever it takes, you'll make it; if you haven't got whatever it takes, you'll find this out. But you'll never really know without trying your best. I wouldn't let any kind of test score determine what I try for. The reality of your own performance in meeting the competition in striving toward your goals is the only real test. I believe this approach gives one the best chances of finally doing what he is best suited for, and this is one of the conditions for a satisfying life."

In statistical terms, however, the correlation is quite substantial between IQ and occupations, when the latter are merely ranked in the order of persons' average judgment of the occupation's prestige. Various studies have shown correlations in the range of .50 to .70. This is sufficiently high that the mean differences between groups of persons in occupations arranged according to a

prestige hierarchy (which is highly related to income) show highly significant differences in IQ or other mental test scores. In general, any two groups which differ in possessing what are perceived as "the good things in life" according to the criteria and values of our society, will be found on the average to differ significantly in IQ. Upward social mobility is related to IQ: the brighter children in a family tend to move up in socioeconomic level and the least bright tend to move down. There are exceptions to the general rule. Those who are born to wealth tend to be less able than those who made it themselves—a quite predictable finding in terms of "regression to the mean." Usually the regression of ability is much greater than the regression of cumulated wealth. The most conspicuous exceptions, however, involve various disadvantaged minorities, whose social and economic positions are different from what one would predict in terms of IQ. For example, Negroes earn less income than whites of comparable IQ, education, family background, and work experience (Duncan et al. 1968). And American Indians, though considerably more impoverished than Negroes in the United States, score higher than Negroes on tests of intelligence and scholastic achievement (Kuttner 1968). Oriental children, who generally score at least as high on IQ as white children, also score considerably higher than would be predicted from their socioeconomic status (SES). This appears especially true of low SES Oriental children, who perform on a par with middle-class white children on nonverbal tests (Lesser, Fifer, and Clark 1965). In predicting scholastic performance in school and in college, however, the evidence indicates that IQ tests and scholastic aptitude tests work with about equal accuracy for all persons from whatever background. In this one respect, at least, the educational system seems to be one of the least discriminatory institutions in our society. For example, there is no evidence that IQ tests predict scholastic performance of Negro children less well than for white children, or that college entrance exams predict college grades less well for Negro than for white students (Jensen 1968b; Stanley and Porter 1967). The predictive validity of such tests could be lowered or changed, of course, by altering the curriculum such that the predictors would no longer be as relevant and other predictors might then become more valid.

When groups are selected from the lower or upper extremes of the IQ distribution, the contrasts are enormous. A classic example is Terman's study of gifted children, selected in elementary school for IQs over 140, which constitutes the upper one percent of the population. These 1,528 children have been systematically followed up to middle age (Terman and Oden 1959). The group as a whole greatly exceeds a random sample of the population on practically every criterion of a successful life, and not just intellectual criteria. On the average the Terman group have markedly greater educational attainments, have higher incomes, engage in more desirable and prestigious occupations, have many more entries in *Who's Who*, have brighter spouses, enjoy better physical and mental health, have a lower suicide rate, a lower mortality rate, a lower divorce rate, and have brighter children (their average IQ is 133). These results should leave no doubt that IQ is related to socially valued criteria.

IS INTELLIGENCE INHERITED?

The evidence on this point is very clear. There is no doubt of a large genetic component in individual differences in IQ. The methodology of determining the heritability of intelligence (or other traits) and the results of the applications of these methods to the study of intelligence have been reviewed in detail elsewhere (Burt 1958; Jensen 1967, 1969a,b). Heritability (H) refers to the proportion of individual differences variance in a measurable trait, like intelligence, that can be attributed to genetic factors. $1\text{-}H$, therefore, is the proportion of variance attributable to non-genetic factors. These non-genetic factors are both biological and psychological. Some substantial proportion of the non-genetic IQ variance is unidentifiable, that is, is due to random environmental effects and to random stochastic biological processes in embryonic development.

The heritability of IQ as estimated from the average of all published studies of the subject is .80, which means that on the average the studies show that 80% of the population variance in IQ is attributable to genetic variation, and 20% to nongenetic factors. The value of .80 is merely an average of many studies which yield H values that range from about .60 to about .90. There is no single true value of the heritability of a trait. Heritability is not a constant, but a population

statistic, and it can vary according to the test used and the particular population sample tested. H will be affected by the range of genetic and environmental variation that exists in the population. It should be noted that all the studies of the heritability of intelligence have been based on European and North American Caucasian populations. The results cannot strictly be generalized to other populations such as American Negroes. We would need to conduct heritability studies within the Negro population if we are to have any certainty that our IQ tests are measuring a genetic component to the same degree in the Negro as in the white population. (Determining H in both populations still would not answer the question whether the group mean IQ difference between Negroes and whites has a genetic component.)

Non-genetic or environmental sources of variance can be analyzed into two major components: variance attributable to differences *between* families in the population and variance attributable to differences *within* families. The sum of the *between* families and the *within* families variances constitutes the total of the nongenetic or environmental variance. Expressed as a proportion, it is $1 - H_B = E$, and, as already pointed out the average value of H_B reported in the literature is .80, making the average value of $E = .20$. The conceptually simplest method for estimating E is to obtain the correlation between identical (monozygotic) twins reared apart (r_{MZA}) in uncorrelated environments (families). $E = 1 - r_{MZA}$. The correlation between identical twins reared together (r_{MZT}) in the same family is used to estimate the *within* families environmental variance $E_W = 1 - r_{MZT}$. The *between* families variance is then $E_B = E - E_W$ or $r_{MZT} - r_{MZA}$. When these formulations are applied to all the relevant twin data reported in the literature, the average values they yield are $E = .20$, $E_B = .12$ and $E_W = .08$ (Jensen 1967). Little, if any, of the E_W is controllable. Some of it is due to prenatal effects related to mother's age, health, accidental perinatal factors, ordinal position among other siblings, etc. In terms of our present knowledge, no prescription could be written for reducing E_W. Some of it, in fact, is almost certainly due to random, stochastic developmental processes in the first weeks after conception, which means that even if we had perfect control over all the identifiable factors usually classified as environmental, genetically identical individuals would still show some differences. The *between* families component, E_B, is probably much more attributable to what we commonly think of as environmental differences in terms of cultural-educational advantages, quality of nutrition, general health care, and the like.

SPECIFIC COMMENTS ON GENETICS AND IQ

To say that IQ tests like the Stanford-Binet "measure present ability, not inborn capacity" is misleading. Surely it measures present ability or performance. But the fact that the heritability (H_B) of the Stanford-Binet IQ is about .80 in English and North American Caucasian populations also means that the test measures "innate capacity," if by this term we mean the individual's genotype for intellectual development. Since H is the proportion of variance in IQs (which are phenotypes) attributable to variance in genotypes, the square root of H represents the correlation between phenotype and genotype, and this correlation is about .90, a very high correlation indeed. (This is the correlation that exists *after* correction for attenuation, that is, test unreliability.) What the evidence on heritability tells us is that we *can*, in fact, estimate a person's genetic standing on intelligence from his score on an IQ test. If the correlation between phenotype and genotype were perfect (i.e., 1.00), a person's test score would, of course, be an exact index of his genetic potential. But since the correlation is only about .90, such statements can only be made on a probabilistic basis.

If education and culturally-derived motivation strongly affect intelligence test performance, then these factors should show up as part of the E variance, mostly E_B, i.e., *between* families environmental variance. Heritability studies, as pointed out, show the E variance to be only about 20% of the total and E_B only about 12% of the total. If *group* differences in IQ are to be explained in terms of educational and motivational factors, and if the heritability of IQ were the same in both groups, it would have to be assumed that all the members of one group differed from the mean of the other group by a *constant amount* in these motivational or other environmental variables. More will be said on this point in the later section on proposed genetic research.

The twin method may actually *underestimate*, rather than overestimate, the heritability of IQ. The reason is that there is considerable evidence that twins are more subject to prenatal stresses and nutritional disadvantages than singletons. This is reflected in the much lower birthweights of twins, the higher infant mortality of twins, and the fact that twins average 6 to 7 points lower in IQ than single-born children. One member of the twin pair is usually prenatally favored over the other, and this is especially true for monozygotic twins, as reflected in their differing birthweights. Birthweight of twins is positively correlated with later IQ (Willerman and Churchill 1967). These prenatal differences, reflected in later IQ differences between the members of twin pairs, are very probably greater for twins than for singletons and therefore suggest a larger component of (prenatal) environmental variance in twins than in singletons. Thus the argument that the twin method of estimating heritability leads to an overestimate and thereby underestimates the environmental component is very weak. A stronger case can be made for just the opposite conclusion. The fact that the estimates of *H* from the twin methods are in close agreement with estimates based on other kinships indicates that the twin estimates are not very deviant in *either* direction. Indeed, it is the consistency of *H* estimates arrived at by different methods that makes them so impressive and reinforces their validity and scientific credibility (see Crow 1969).

Cultural and educational differences are probably the most important *non-biological* sources of individual differences in intelligence, but they are not necessarily the most important *non-genetic* source of differences. It is likely that prenatal and nutritional factors are at least as important sources of variance as social-psychological factors. The sociological emphasis on the non-biological aspects of the environment has resulted in a relative neglect of probably important nutritional factors and maternal factors (age, health, diet, number of births, spacing of births, etc.) which can affect the prenatal and early childhood development of the individual.

In reply to suggestions that our national IQ may be declining due to the possibility that the least able segment of the population is reproducing at a faster rate than the most able segment, some writers draw the familiar analogy between intelligence and physical stature. Both IQ and height are polygenetic traits and the same quantitative genetic model can be applied to both and can predict the various kinship correlations for IQ and height about equally well. It is also known that height, like intelligence, shows a positive correlation with socioeconomic status. Thus, if poor people have larger families than the well-to-do, we should expect the average height of the population to decrease over a number of generations. Exactly the same line of reasoning applies also in the case of intelligence. To counter this pessimistic prediction, it has been noted that despite what we should predict from simple genetic principles, the mean height of the population not only has not decreased in the past 200 years or so, but has in fact *increased* by a very significant amount. The increase, it is assumed, is due to environmental factors such as improved nutrition. And the implication is, of course, that intelligence, too, will increase over generations because of improvements in the environmental factors relevant to intellectual development. I believe this line of argument is weak and can lead to an unwarranted complacency about a possibly serious social trend.

First of all, Carter (1962) and Tanner (1965, 1968) have pointed out that much if not all of the increase in adult height in the past 200 years can be attributed to genetic factors, namely, the outbreeding effect. Increase in height is closely associated with the increase in the population's mobility. The offspring of parents from different Swiss villages, for example, are taller than the offspring of parents born in the same village. This outbreeding effect, or hybrid vigor, tends to saturate or level off in the population in a few generations, as has already occurred with respect to height in the United States. Nutritional factors have their greatest effect on *rate* of growth rather than on final adult height. In World War I men reached their full adult height at age 26; today they attain their full height at 18 or 19.

Although it is true that height is positively correlated with socioeconomic status (SES) and that low SES families are larger than high SES families, these facts alone are not sufficient to warrant the prediction that the mean height of the population should decline. It would have to be shown that the same *numbers* of low SES persons as high SES persons have offspring. When this point was investigated for intelligence, it was found that persons of below average IQ have larger families than persons of above average IQ, but that fewer of the below average ever marry or have any

children at all (Higgins and Reed, 1962; Bajema 1963, 1966). The net result is a balance between the low and high IQ groups in the number of offspring they produce. This finding holds only for the white population of the US of a generation ago. No studies of this type have been conducted in the US Negro population. Since the bases for marriage and mate selection may be quite different in various subcultures, the results of investigation of this problem in one group cannot be generalized to other population groups with any confidence. The analogy with height is not convincing, since we have established only a negative correlation between height and family size, but have not taken into account the relative proportion of short and tall persons who never marry or produce offspring. Since we know there is selective mating for height in our population (that is, taller persons are viewed as more desirable) it is likely that fewer short persons marry or reproduce and that therefore a similar equilibrium between reproductive rates of short and tall persons exists as in the case of low and high intelligence. As I have noted elsewhere (Jensen 1968a, 1969a), certain statistics raise the question of whether Negro intelligence is declining relative to white intelligence as a result of more extreme differential birthrates in lower and upper social classes among Negroes than among whites. Negro middle- and upper-class families have fewer children than their white counterparts, while Negro lower-class families have more. In 1960, Negro women married to professional or technical workers had only 1.9 children as compared with 2.4 for white women in the same circumstances. Negro women of ages 35 to 44 who were married to unskilled workers had 4.7 children compared with 3.8 for non-Negro women in the same situation, and Negro women with incomes below $2000 per year averaged 5.3 children (Moynihan 1966). This could mean that the least able segment of the Negro population is reproducing most rapidly, a condition that could alone produce and increase a genetic difference between the Negro and white populations in a few generations. The possible genetic and social implications of these trends have not yet come under investigation and there are no data at present which would warrant complacency about this important question.

Can genetic changes in a population take place only very slowly, so that selective pressures acting over several generations would be of negligible consequence? The answer, of course, depends largely on the degree of selective pressure. We already know enough to permit fairly accurate estimates of genetic trends given certain criteria of selection. If selection were extremely rigorous, an enormous shift in the population mean would be possible, as can be inferred from the average IQ of the offspring of the Terman gifted group. The Terman subjects were selected for Stanford-Binet IQs of 140 and above; they had a mean of 152. There was no selection of their spouses, except by the normal assortative mating that occurs for intelligence in our society (i.e., a correlation of .5 to .6 between spouses' IQs). The offspring of the Terman gifted had an average IQ of 133 (Terman and Oden 1959). This is more than two standard deviations above the mean IQ of children born to a random sample of the population. There is a regression from the selected parent generation toward the general population mean, but the regression happens only once, and the offspring of the selected parents will in turn have offspring without further regression, provided, of course, they do not mate outside the group of offspring from the selected parents. Rats have been bred for maze learning ability and it has generally required from six to nine generations of selection to produce two strains of rats whose distributions of maze learning scores are completely non overlapping.

IS RACE A VARIABLE?

One of the easiest ways of avoiding the issue of race differences in intelligence is to make the claim that there is no such thing as race and therefore it is not a variable that can be related to any other variables. Thus, proponents of this view would claim that the concept of race is merely a myth, not a phenomenon that can be subjected to scientific study. This is, of course, utter nonsense. But it will pay to clarify the concept of race as it figures in comparative studies of intelligence.

There are two general definitions of race: the social and the biological (or genetic). Both are arbitrary, but this need not mean they are unreliable or lacking in precision. Although most of the studies of racial differences in intelligence are based on social definitions of race, it should be

noted that there is usually a high correlation between the social and the biological definitions, and it is most unlikely that the results of the research would be very different if the investigators had used biological rather than social criteria of race in selecting groups for comparisons.

The social criteria of race are simple: they are the ethnic labels people use to describe themselves and the more obvious physical characteristics such as skin color, hair texture, facial features, and so on, by which persons roughly determine one another's "race." Admitted, the social definition is crude. It does not take account of "borderline" or ambiguous cases that are hard to categorize and which make for some unreliability in classification, and it does not take account of the fact that there are no pure racial types—and especially in the case of American Negroes there is considerable racial admixture. Almost no American Negroes are of pure African descent; most have from 5% to 90% Caucasian genes, the average degree of admixture now being between 20% and 30%. Thus there is great genetic diversity *within* socially defined racial groups.

Does this make the social definition of race useless as a variable? No. In the first place, there is undoubtedly a high correlation between social and biological classification. That is to say, if one were to sort school children, for example, into three socially defined racial groups, Negro, Oriental, and Caucasian, one would find a very high concordance of classification if he used strict biological criteria based on the frequencies of blood groups, anthropometric measures, and other genetic polymorphisms. What one would not have obtained from the crude social classification is degrees of racial admixture. In other words, the major racial categories would be much the same whether constituted by social judgments or strict biological criteria. But if we wanted to go beyond this crude system of classification to make more refined differentiations, we would have to resort to biological criteria. Social judgments of degrees of racial admixture are quite unreliable. The broad categories, however, *are* reliable. They also qualify as variables in the sense that they show significant correlations with other variables such as IQ and scholastic performance. This is not to say that such correlations by themselves tell us anything about a biological or genetic basis for the correlation, which might be due to other environmental, social-class and cultural variables related to the socially defined racial classification. If the crucial variables in IQ differences are not racial classification per se, but other correlated environmental factors, then, at least in theory, one should be able to reduce the racial correlation with IQ to zero by partialling out the truly causal factors that are only incidentally correlated with both race and IQ. So far no one has succeeded in doing this as regards Negro-white comparisons. Every combination of environmental variables that anyone has partialled out has always left behind some significant correlation between race (socially defined) and IQ (Shuey 1966). One can always claim that all the relevant environmental variables were not taken into account. This is a real weakness of such studies and they can be legitimately criticized on this score. It is largely for this reason that our understanding of racial differences will not be greatly advanced until more refined criteria of race based on biological criteria are employed. Specific proposals are made in a later section.

It is strange that those who claim that there are no genetic racial differences in ability are often the most critical of studies that have employed the social criterion of race rather than more rigorous genetic criteria. If the observed IQ differences are due only to social factors, then the social definition of race should be quite adequate, and, in fact, should be the only appropriate definition. If it is then argued that the two socially defined racial groups being compared are not "pure" and that each group contains some genetic admixture of the other, it can only mean that the biological racial aspects of the observed IQ differences has been *underestimated* by comparing socially defined racial groups.

The biological definition of race is based on gene frequencies. Races are breeding groups which differ in the frequencies of one or more genes. A breeding group is one in which there is a higher proportion of matings among members of the group than of matings in which one member of the pair is from outside the group. Breeding groups result from relative degrees of geographical, racial, and cultural isolation of different population groups. The definition of race by these criteria is arbitrary only in the sense that differences in gene frequencies is a continuous variable, and where one wishes to draw the lines as criteria for classification purposes is not dictated by nature but by the taxonomic considerations of the investigator. Rather than thinking in terms of races, we should think in terms of groups with different gene frequencies. The question we would ask is

whether various groups differing in gene frequencies also differ in IQ, other things being, in effect, equal. The major races are simply breeding populations that have a relatively high degree of inbreeding and differ from one another in the relative frequencies of many genes. They differ in so many known gene frequencies, in fact, that it seems highly improbable that they would not also differ in the frequencies of genes related to behavioral traits such as intelligence.

A major block to clear thinking about race is to think of it as a kind of Platonic essence, independent of any particular population group. General statements about the mental abilities of the "white race," the "black race," "the yellow race," and so on, make no sense in terms of any studies that have yet been done or that seem at all feasible for the future. Strictly speaking, to ask if there are race differences in any characteristic is scientifically meaningless if what we mean by race is not clearly specified. All we can do is study samples selected from certain specified populations. These samples cannot be regarded as representative of some Platonic racial groups. They are merely representative (if properly selected) of the clearly specified population group from which they are selected.

We could ask, for example, whether a population subgroup that differs from the general population in its average response to the educational and occupational requirements of our society differs in its gene pool from other population subgroups which are more successful, and if so, are some of the genetic differences related to ability factors with high heritability?

Population subgroups which have immigrated are not necessarily representative of their native parent populations. Studies of racial or national groups in the United States, therefore, cannot be generalized abroad, and the reverse is also true. This does not mean, however, that meaningful comparative studies of various population subgroups within the United States are not feasible.

The notion that there are no genetic mental ability differences among population subgroups that differ in many other gene frequencies is, in principle, hard to defend. Populations that have been widely separated geographically or socially for many centuries and which have been exposed to climatic and cultural conditions that exert different selective pressures are almost certain to differ genetically in many ways. And, in fact, they do. Nearly every anatomical and physiological system studied has shown race differences. It is not at all necessary to invoke the factor of differential selective pressures to validate or explain some of these genetic differences, many of which confer no discernible advantage or disadvantage to survival or adaptation in any particular environment. A chemical substance, phenylthiocarbamide (PTC), is one illustration. To some persons PTC is completely tasteless; to others it has a very unpleasant bitter taste. Whether a person is a taster is determined by a single gene. This gene has markedly different frequencies in different racial groups. No one knows why this should be. Similarly, blood types have markedly different distributions in various racial groups, although it is not at all clear that one blood type is more advantageous than another in any given environment. In short, genetic diversity is the rule; genetic uniformity is the rare exception. By definition the gene pools of racial groups differ, and it is not at all an unreasonable hypothesis that genetic factors that condition behavioral development also differ.

Biological evolution generally is a slow process, but genetic changes with respect to particular traits can occur relatively fast in response to selective pressures in the environment. In any case, biological evolution, whatever its rate, has resulted in marked genetic differentiation of human populations. Concerning the one standard deviation average IQ difference between Negro and white American populations, one writer stated, "A review of present knowledge on interracial divergence in man makes it unlikely that a difference as large as the observed one is genetic." This hardly seems tenable in view of the fact that other traits show even greater racial differences than are found for intelligence. Height, like intelligence, is a polygenically inherited characteristic and is probably less subject to selective pressures than intelligence, and yet we find racial (and even national or regional) differences of more than one standard deviation. In fact, two racial subgroups on the African continent, the Pygmies and the Watusi, differ in height by five to six standard deviations. Obviously biological evolution has, in fact, been sufficient to create marked differences in genetic characteristics.

It is hard to imagine that there have not been different selection pressures for different abilities in various cultures and that these pressures would be as great for intelligence as for many physical

characteristics which are known to differ genetically among racial groups. Individual differences in the abilities most relevant to a particular culture are highly visible characteristics and if they have consequences for the individual's status in the social hierarchy or the culture's system of rewards they will be traits subject to the genetic effects of sexual selection and assortative mating. If a trait is not very relevant to the demands of a particular culture it will not become highly visible, it will not be a basis for selective mating, and its genetic basis will not be systematically affected by pressures in the social environment.

Selective mating refers to the fact that certain characteristics are viewed as desirable in mate selection by virtually all members of the breeding population. The usual consequence is that those standing higher on the desired trait will have greater opportunities for mating and reproduction while those at the lowest end of the distribution on the trait in question will be least likely to find a mate and to leave progeny. The net effect is to boost the mean value of the trait in the population. Assortative mating refers to the fact that like tends to marry (or mate with) like. It is sometimes an inevitable consequence of selective mating with respect to generally desirable traits, but also holds for traits which are merely subject to various individual preferences. It is noteworthy that of all measurable human characteristics the one with the highest coefficient of assortative mating (i.e., the correlation between mates) is intelligence. The correlation between spouses' IQs, for example, are around .5 to .6 in various studies, as contrasted with a correlation of .3 for height and of zero for fingerprints. The high degree of assortative mating for intelligence means that it is highly subject to genetic change through social influences. For example, the variance of the IQ distribution in the population would be reduced by approximately 20% if there were no assortative mating for just one generation. Assortative mating increases the variance of the characteristic in the population, and if there is selective mating (as well as assortative) for the characteristic, the individuals at the lower (least desirable) end of the distribution will be least likely to reproduce. The net effect is to raise the average of the population on the trait in question. Such trends have probably taken place with respect to different traits in different societies for many centuries. While sexual selection may be capricious and non-adaptive with respect to many physical characteristics (e.g., various societies have different criteria of beauty), selection is not likely to be capricious with respect to those abilities which are salient in the competition in a given society. There has probably been quite strong and consistent selection for different patterns of ability in different cultures. A high degree of genetic adaptation to the demands of one environment might not constitute optimal adaptive capabilities to the demands of another, quite different, environment. As stated by Spuhler and Lindzey (1967:413) in their chapter on the behavior-genetics of race difference:

> it seems to us surprising that one would accept present findings in regard to the existence of genetic, anatomical, physiological, and epidemiological differences between races and still expect to find *no* meaningful differences in behavior between races.

They continue to point out that there are

> enormous discrepancies between races in the efficiency with which culture is transmitted (for example, the difference between literate and nonliterate societies). Some of these differences are closely associated with race differences, have existed for many thousands of years, and presumably have been accompanied by very different selection pressures in regard to characters potentially relevant to culture transmission, such as 'intelligence.'

Thus, it seems highly improbable that there have been no markedly differing selective pressures on different subpopulations even within the United States. The selective pressures on Negroes must have been very different from those in European immigrant populations. The history of slavery suggests quite extreme selective factors, involving even the deliberate breeding of slaves for certain characteristics which were irrelevant or perhaps even negatively correlated with intellectual prowess. It would be surprising indeed if more than 300 years of slavery did not have some genetic consequences. But since the possible nature of these consequences is highly speculative and cannot be accurately inferred from historical accounts, this retrospective approach to the study of racial differences is too unreliable to be of much real scientific value. Direct genetical studies of present population groups can provide the only really satisfactory basis for the scientific study of genetic differences in abilities.

ARE THERE RACIAL DIFFERENCES IN IQ?

In the United States persons classed as Negro by the common social criteria obtain scores on the average about one standard deviation (i.e., 15 IQ points on most standard intelligence tests) below the average for the white population. One standard deviation is an *average* difference, and it is known that the magnitude of Negro-white differences varies according to the ages of the groups compared, their socioeconomic status, and especially their geographical location in the United States. Various tests differ, on the average, relatively little. In general, Negroes do slightly better on verbal tests than on non-verbal tests. They do most poorly on tests of spatial ability, abstract reasoning and problem solving (Shuey 1966; Tyler 1965). Tests of scholastic achievement also show about one standard deviation difference, and this difference appears to be fairly constant from first grade through twelfth grade, judging from the massive data of the Coleman study (1966). The IQ difference of 1 SD, also, is fairly stable over the age range from about 5 years to adulthood, although some studies have shown a tendency for a slight increase in the difference between 5 and 18 years of age. Another point that has been suggested, but which requires much more systematic investigation before any firm conclusions can be reached, is that there is a larger sex difference in IQs for Negroes than for whites (Bronfenbrenner 1967). The presumed difference favors the females. The point is especially worthy of research because, if true, it would have considerable social and educational consequences, which would be especially evident in the upper tail of the IQ distribution. For example, if girls are a few IQ points higher than boys, on the average, one should expect a greatly disproportionate number of Negro girls to qualify, as compared with boys, in any selection based on cut-off scores well above the mean, such as selection for college. Assuming a general mean of 85, an SD of 15, and a normal distribution, a 5 point IQ difference between Negro boys and girls and a college selection cut-off score of 115, for example, we would expect the number of qualified girls to boys to be approximately in the ratio of 2 to 1.

A statistic which has been much less studied than the mean difference is the standard deviation (SD), that is, the measure of dispersion of scores within the distribution.

Most studies agree in finding a smaller SD in Negro than in white IQs. The single largest normative study of Stanford-Binet IQs in a Negro population, for example, found an SD of 12.4 as compared with 16.4 in the white normative sample (Kennedy, Van de Riet, and White 1963). This study is based on a large sample of school children in five Southeastern states and therefore may not be representative of the Negro population in other regions of the US. In general, however, most studies of Negro intelligence have found a smaller standard deviation than the SD of 15 or 16 generally found in white samples. The point is of some consequence in considering the relative merits of the opposing hypothesis relating to the *causes* of the observed average IQ difference between Negroes and whites, namely, the hypothesis of genetic equality versus the hypothesis of genetic differences. If the distribution of IQs in the Negro population does, in fact, have a smaller SD than in the white population, and if we hypothesize no genetic differences between the two populations, we must conclude that there is less variance due to environmental differences within the Negro group than within the white group. Since the genetic variance is hypothesized to be exactly the same in both groups, the difference in the variances (i.e., the square of the SD) of the groups must be all environmental variance. Thus, if the total variance of Negro IQs is less than of white IQs, the genetic equality hypothesis is forced to predict a higher heritability of IQ in the Negro population than in the white; that is to say, more of the variance in Negro IQs would have to be due to genetic factors. If a study of the heritability of IQ in the Negro population yielded a heritability coefficient equal to or less than that found in the white population, this finding would contradict the genetic equality hypothesis, at least as regards the equality of genetic variance in the two populations.

Let us take another look at the Kennedy et al. (1963) data in this connection, to see how the hypothesis of genetic equality of variances comes out for this one set of data comparing the distribution of Negro IQs with the distribution of the white population sample on which are based the norms for the Stanford-Binet Intelligence Test. It will be recalled that the SDs for Negroes and whites were 12.4 and 16.4, respectively. The variances are thus $(12.4)^2 = 153.76$ and $(16.4)^2 =$

268.96. Now, the best estimate of the heritability and Stanford-Binet IQs in white population samples similar to that on which the Stanford-Binet was standardized is .80 (Jensen 1969a). This means that 80% of the variance of the white IQ distribution is *genetic* variance: thus, .80 x 268.96 = 215.17 is the white genetic IQ variance. But this is still greater than the *total* Negro IQ variance. The heritability of IQ in the white group would have to be assumed to be .57 for the white *genetic* variance to equal the *total* IQ variance of the Negro group, and surely some of this total variance is non-genetic. Furthermore, no reported study of the heritability of Stanford-Binet IQs is as low as .57. Thus, a hypothesis of genetic equality with respect to variances leads to highly untenable conclusions when applied to the data of Kennedy et al. (1963). By any canon of statistical and logical reasoning one is forced to reject the hypothesis that the distributions of genotypes for intelligence are equivalent in these two samples. By assuming genetic equivalence, one simply cannot make any sense out of the available data. This is not to say that one cannot question the data with respect to every parameter that is involved in this line of reasoning.[1] But if one accepts the validity of the heritability estimates in the white population and the SDs given by Kennedy et al., it logically follows that a genetic equivalence hypothesis is untenable. It is, of course, statistically unwarranted to generalize this conclusion beyond the populations sampled in the study by Kennedy et al. The cause of the lesser variance of IQ in the Negro group is not known. One can only speculate and suggest hypotheses. From the evidence on the white population, for example, we know that some 15% to 20% of the total variance is attributable to assortative mating for intelligence; if the correlation between mates' IQs was markedly reduced, the white IQ variance would be substantially reduced. (Variance due to assortative mating is all *genetic* variance.) Also, the covariance of heredity and environment (i.e., there is some correlation between children's genotypes for intellectual development and the quality of the environment in which they are reared) constitutes some 5% to 10% of the total IQ variance in the white population. If environments were more similar, there would be less covariance and this source of variance would be diminished in the total. We could find out if these factors or others, or some combination of factors, are responsible for the lesser variance in the Negro population only by carrying out complex heritability studies in the Negro population.

A point that should be stressed is the fact that neither the white nor the Negro population, by common social classification, is genetically homogeneous. It has already been noted that the American Negro is not of pure African ancestry but has, on the average, an admixture of 20% to 30% Caucasian genes, varying from less than 5% in some regions of the country to 40% or 50% in others (Reed 1969). The white population contains many different subgroups which most probably differ genetically in potential for intellectual development. To point to one particular subgroup of one socially defined racial population as being higher or lower in IQ than some subgroup in another racial population proves nothing other than the fact that there exists an overlap between the racial groups. The fact that relatively large mean IQ differences are found between certain subgroups within the same race does not mean that these differences must be entirely of environmental origin and that therefore racial differences of similar magnitude must also be entirely attributable to environment.

Finally, it should be noted that IQ tests are taken by individuals. There is no such thing as measuring the IQ of a group as a group. Individuals' IQs are obtained as individuals. The basis on which individuals may be grouped is a separate issue, depending upon the purpose of the investigator. When test scores are grouped according to some criteria of racial classification, we find mean differences between the groups. If we group test scores by some criteria of socioeconomic status, we find mean differences between the groups. Conversely, if we group persons by levels of IQ, we find the groups differ in their proportions of persons of different races and social classes.

ARE RACE DIFFERENCES IMPORTANT?

There is, of course, nothing *inherently* important about anything. Race differences in intelligence are important only if people think these differences, or their consequences, are important. It so happens that in our society great importance is given to these differences and their

importance is acknowledged in many official public policies. Racial inequality in educational and occupational performance, and in the social and economic rewards correlated therewith, is today clearly one of the uppermost concerns of our nation.

Most persons are not concerned with those racial characteristics that are patently irrelevant to performance. The real concern results from the observed correlation between racial classification and educational and occupational performance. Persons who feel concerned about these observed differences demand an explanation for the differences. It is apparently a strongly ingrained human characteristic to need to understand what one perceives as a problem, and to ask for answers. People inevitably demand explanations about things that concern them. There is no getting around that. We have no choice in the matter. Explanations there will be.

But we do have a choice of essentially two paths in seeking explanations of intelligence differences among racial groups. On the one hand, we can simply *decree* an explanation based on prejudice, or popular beliefs, or moral convictions, or one or another social or political ideology, or on what we might think it is best for society to believe. This is the path of propaganda. Or, on the other hand, we can follow the path of science and investigate the problem in the same way that any other phenomena would be subjected to scientific study. There is nothing to compel us to one path or the other. This is a matter of personal preference and values. And since persons differ markedly in their preferences and values, we will inevitably see both of these paths being followed for quite some time. My own preference is for a scientific approach to the study of these phenomena. It is certainly the more interesting and challenging intellectually. And our experience tells us that the scientific approach, by and large, leads to more reliable knowledge of natural phenomena than any other method that man has yet devised. If solutions to educational problems depend upon recognizing certain psychological realities in the same sense that, say, building a workable spaceship depends upon recognizing certain physical realities, then surely we will stand a better chance of improving education for all children by choosing the path of scientific investigation. In facing the issue of race differences in abilities we should heed the statement of John Stuart Mill:

> If there are some subjects on which the results obtained have finally received the unanimous assent of all who have attended to the proof, and others on which mankind have not yet been equally successful; on which the most sagacious minds have occupied themselves from the earliest date, and have never succeeded in establishing any considerable body of truths, so as to be beyond denial or doubt; it is by generalizing the methods successfully followed in the former enquiries, and adapting them to the latter, that we may hope to remove this blot on the face of science.

Once we subscribe to a scientific approach, we are obligated to act accordingly. This means, for one thing, that we entertain alternative hypotheses. To entertain an hypothesis means not just to pay lip service to it or to acknowledge its possible merit and let it go at that. It means to put it into a testable form, to perform the test, and report the results with information as to the degree of statistical confidence with which the hypothesis in question can be accepted or rejected. If we can practice what is called "strong inference," so much the better. Strong inference consists of formulating opposing hypotheses and pitting them against one another by actually testing the contradictory predictions that follow from them. This is the way of science. How much of our educational research, we may ask, has taken this form? How much of the research that we see catalogued in the already gargantuan ERIC bibliography on the causes of the educational handicaps of children called culturally disadvantaged has followed this path? The only sensible conclusion one can draw from a perusal of this evidence is that the key question in everyone's mind about racial differences in ability—are they genetic?—has, in effect, been ruled out as a serious alternative hypothesis in the search for the causal factors involved in inequalities of educational performance. Sundry environmental hypotheses are considered, but rarely, if ever, are alternative genetic hypotheses suggested. If a genetic hypothesis is mentioned, it is usually for the sake of dismissing it out of hand or to point out why it would be impossible to test the hypothesis in any case. Often, more intellectual ingenuity is expended in trying to find reasons why a particular genetic hypothesis could not be tested than in trying to discover a way of formulating the hypothesis so that it could be put to a test. The emotional need to believe that genetic factors

are unimportant in individual or group differences in ability can be seen in many statements by dedicated workers in those fields of psychology and education most allied to the problems of children called disadvantaged. For example, Dr. Bettye Caldwell, a prominent worker in compensatory and early childhood education has noted:

> Most of us in enrichment . . . efforts—no matter how much lip service we pay to the genetic potential of the child—are passionate believers in the plasticity of the human organism. We need desperately to believe that we are all born equalizable. With any failure to demonstrate the effectiveness of compensatory experiences offered to children of any give age, one is entitled to conclude parsimoniouslv that perhaps the enrichment was not offered at the proper time [Quoted by Baratz and Baratz 1969].

But genetic factors in rate of development are never considered as a possible part of the explanation. It is important not to evaluate persons in terms of group membership if we are to insure equality of opportunity and social justice. All persons should be treated as individuals in terms of their own merits if our aim is to maximize opportunities for every person to develop his abilities to their fullest capacity in accord with his own interests and drives. But the result of *individual* selection (for higher education, better jobs, etc.) makes it inevitable that there will be unequal representation of the parent populations in any subgroup that might be selected whenever there are average differences between parent populations.

Many questions about the means of guaranteeing equality of educational opportunity are still moral and political issues at present. When there is no compelling body of scientific evidence on which policy decisions can be based, such decisions must be avowedly made in terms of one's personal social philosophy and concepts of morality. Many goals of public policy must be decided in terms of values. The results of research are of greatest use to the technology of achieving the value-directed goals of society. The decision to put a man on the moon was not a scientific decision, but once the decision was made the application of scientific knowledge was necessary to achieve this goal. A similar analogy holds for the attainment of educational goals.

CAN RACE DIFFERENCES BE RESEARCHED?

It is sometimes argued that even though it is not unreasonable to hypothesize genetic racial differences in mental ability, we cannot know the direction or magnitude of such genetic differences and the problem is much too difficult and complex to yield to scientific investigation. Therefore, the argument often continues, we should go on pretending as though there is no question of genetic differences, as was officially stated by the US Office of Education in 1966: "It is a demonstrable fact that the talent pool in any one ethnic group is substantially the same as that in any other ethnic group."

First, we will never know to what extent research can yield answers on a subject unless we at least try our best to do the research. It is doubtful that any major scientific advances could have been made in any field if it were decided beforehand that the problems could not be researched. I cannot agree that a scientific approach should be restrictied to only the easy problems. If all the necessary methodology for studying the genetics of race differences in psychological characteristics is not yet sufficiently developed, this should not be surprising, since so little effort has been made thus far. The methodology of a field of inquiry does not grow in a vacuum. Scientists do not *first* develop a complete methodology for the investigation of a complex area and then apply it all at once to get the final answers. An appropriate methodology evolves as a result of grappling with difficult problems in the spirit of scientific research. Darwin's theory of evolution did not begin with a fully developed methodology adequate to prove the theory, nor did the theory of the inheritance of acquired characteristics—a theory which was later disproved after the development of an adequate methodology, a methodology which would not have developed in the absence of attempts to research this theory. No one would have been inclined to invent the necessary research methods in the absence of the problems these methods were needed to solve. One critic states "The scientific problem [of genetic race differences in ability] itself seems of dubious validity, if one considers how great are the difficulties . . ., at least on the basis of present techniques." The same statement could have been made about research on the theory of evolution,

the atomic theory, the gene theory, and so on. We do not expect any single study or experiment to reduce all the uncertainty about a complex subject to absolute zero in one bold stroke! But as in dealing scientifically with most other complex phenomena, we should not regard ourselves as so intellectually impotent as to be unable to gradually chip away at the heredity-environment uncertainty with whatever tools that scientists can muster or devise with their present knowledge and ingenuity.

What are some of the thinking blocks in this area? One is the frequent failure to distinguish between raw facts, on the one hand, and inference from the facts in terms of some hypothesis, on the other. The Society for the Psychological Study of Social Issues (SPSSI), for example, in a press release (May 2, 1969) criticizing my article in the *Harvard Educational Review* (Jensen 1969a), stated, "There is no *direct* [italics mine] evidence that supports the view that there is an innate difference between members of different racial groups." Of course there is not *direct* evidence, nor can there be direct evidence if by "direct" we mean evidence that is immediately palpable to our physical senses. The gradual disappearance of ships over the horizon is not *direct* evidence of anything, but it can be interpreted in terms of the hypothesis that the earth is round. It would be harder to explain if we hypothesized that the earth is flat. So even as relatively simple an hypothesis as that the world is round cannot be proved by direct evidence, but depends upon logical inference from diverse lines of evidence. If all that was needed was direct evidence, even a monkey would know that the world is round, in the same sense that it knows that a lemon is sour. The substantiation of an hypothesis in science depends upon *objective* evidence but does not necessarily depend upon direct evidence alone.

Another inhibition to thought on this topic is the notion that before research can yield any answers, the environment must be absolutely equal for all groups involved in comparisons. The SPSSI statement went so far as to say that ". . . a more accurate understanding of the contribution of heredity to intelligence will be possible only when social conditions for all races are equal and when this situation has existed for several generations." Since no operationally testable meaning is given to "equal" social conditions, such a statement, if taken seriously, would completely preclude the possibility of researching this important question, not just for several generations, but indefinitely. Actually, large environmental differences between racial groups can be revealing when the environmental ratings are positively correlated with IQ or scholastic performance *within* the groups but show a negative correlation *between* the groups. If group A on the average has a poor environment in terms of variables claimed to be important to intellectual development and group B has a good environment, and if group A performs better than group B on intelligence tests which are appropriate to the experience of both groups, this is evidence that some factors other than the measured environmental variables are involved in the relatively higher intellectual performance of group A as compared with group B. If environmental factors cannot be found that will account for the difference, it is presumptive evidence in favor of the genetic hypothesis. Genetical tests of the hypothesis are preferable, of course. (These are discussed in a later section). But what one also looks for are consistencies among various lines of evidence, especially lines of evidence that lead to opposite predictions from different hypotheses.

Many investigators now would question the view that the lack of early stimulation in the preschool years can be counted among the chief causes of the poorer IQ performance of Negro children, since when children are grouped in several categories according to their parents' socioeconomic status, the Negro children in the highest SES category still score two to three IQ points below white children in the lowest SES level (Shuey 1966). Thus, what we generally think of as a reasonably good environment is apparently not sufficient to equalize the performance of Negro and white groups.

Such findings lead to hypothesizing increasingly subtle and hard to measure environmental effects. But it should be recognized that at present most of the environmentally "damaging" effects that are assumed to be accountable for performance differences are hypothetical and not factual. Poor self-concept and alienation are among the currently prevailing explanations, but what has not yet been satisfactorily explained is why such general motivational dispositions should affect some cognitive abilities so much more than others. Performance is not uniformly low on all tasks, by any means. There are distinct high and low points in the profile of various abilities in

77-942 O - 72 - 16

different ethnic groups (Stodolsky and Lesser 1967), and no one has yet attempted to explain how such profile differences, which are invariant across social classes, could come about as a result of differences in generalized attitudes and motivation in the test situation.

Finally, unnecessary difficulties arise when we allow the scientific question to become mixed up with its possible educational, social, and political implications. The scientific question and its solution should *not* be allowed to get mixed up with the social-political aspects of the problem, for when it does we are less able to think clearly about either set of questions. The question of whether there are or are not genetic racial differences in intelligence is independent of any questions of its implications, whatever they may be. But I would say that the scientific question should have priority and the answer should be sought through scientific means. For although the answer might have educational and social implications, and there are indeed grave educational and social problems that need to be solved, we must first understand the causes of problems if we are to do anything effectively toward solving them. Gaining this knowledge is a scientific task. As it is accomplished, we are then in a better position to consider alternative courses of action and evaluate their feasibility and desirability in terms of society's values and goals. This moves the problem into the realm of public policy, where all the answers cannot be scientifically derived. But policy cannot be wisely or effectively formulated unless it is informed by the facts. No matter how well-intentioned it may seem to be, it can only be less effective and less beneficial if it is based on false premises or in contradiction of reality.

GENETIC RESEARCH TO REDUCE THE HEREDITY-ENVIRONMENT UNCERTAINTY

Today there is virtually no uncertainty among those who have attended to the evidence that individual variation in intelligence is predominantly conditioned by genetic factors and that environmental factors account for a lesser proportion of the phenotypic variance. One can point to variations among studies that have estimated the heritability of intelligence. Such variations in estimates of the proportion of variance attributable to genetic factors are to be expected in view of the great variety of populations sampled and the differences among the variety of tests of mental ability that have been used. Despite these expected variations in heritability estimates, it is important to note that no major study contradicts the conclusion that heredity contributes something more than twice as much to the variance in IQ as environment in *white* European and American populations. (We do not have good heritability data on other populations.)

The term "heredity-environment uncertainty" refers mainly to the question of race differences in intelligence. The answer to this question is still in the realm of uncertainty in terms of the normal scientific meaning of this word. *Absolute* certainty is never attained in an empirical science. Absolute certainty can be had only in pure mathematics, the certainty of which rests upon the fact that pure mathematics is, as Bertrand Russell pointed out, just one vast tautology. Empirical science deals in probability statements, and "certainty" refers to a high degree of probability that a proposition is "true," meaning that certain objective consequences can be predicted from the proposition with a stated probability. A decisive increase in this probability with respect to any given scientific proposition rarely results from a single experiment or discovery. I take exception to the impression that might be given by some writers that unless a scientific study can be perfect and 100% certain, we cannot know anything. This is not how scientific knowledge advances. We do not devise perfect methods or obtain complete answers on the first try. Certainty, in the sense of probability, is generally increased very incrementally in science. Research aims to add reliable increments to statements of probability.

This we must continue to do with respect to the question of genetic race differences in intelligence. It is still an open question by all reasonable scientific standards. The existing evidence is in all cases sufficiently ambiguous, due largely to the confounding of racial and environmental factors, as not to permit statements with a sufficiently high probability such that all reasonable and qualified persons viewing the evidence will agree that it is conclusive. The issue of genetic race differences may be likened to theories of the moon's craters—whether they were caused by volcanic eruptions or by the impact of meteors. All the evidence obtainable by astronomers could

support either interpretation, and different scientists could argue for one theory or the other. A substantial increment could be subtracted from this uncertainty only by obtaining new evidence not obtainable through telescopic study, namely, directly obtaining and analyzing material from the surface of the moon.

I believe that, similarly, the heredity-environment uncertainty about race differences in IQ will be substantially reduced only by obtaining new evidence—new *kinds* of evidence. Exclusive reliance on anthropological, sociological, and psychological evidence would probably not substantially advance our knowledge. I believe that application of the methods of biometrical genetics (also called population genetics or quantitative genetics) to the question of race differences will substantially reduce our uncertainty.

Someone suggested that the only way one could prove race differences in intelligence would be to dye one member of a pair of white identical twins black and adopt it out to a Negro family while the co-twin is reared by a white family. How much difference would it make in their IQs? Better yet is the suggestion of Professor Arthur Stinchcombe (1969): find pairs of identical twins in which one member of each pair is Negro and one is white, separate them at birth and rear them in Negro and white families and see how their IQ differences compare with those found for twins where both are of the same race! These suggestions sound ridiculous; one is unfeasible and the other is impossible. Yet as conceptual experiments they are good, because they suggest the necessary ingredients of the information we must obtain to reduce the heredity-environment uncertainty. Both examples rightly recognize skin color (and, by implication, other visible racial features) as a part of the individual's environment. They are based on comparing genetically equivalent persons reared in different environments. Another possibility consists of rearing genetically and racially different persons in essentially similar environments—including the factor of skin color, etc. Is such a study possible? Yes.

Geneticists already know the frequencies of a large number of genetically independent blood groups in European and African populations. On the basis of such data, it is entirely possible to determine the proportion of Caucasian genes in a population sample of Negroes, socially defined. Furthermore, it should be possible by the same means to classify individuals on a probabilistic basis in terms of their relative proportions of African and Caucasian genes. Since the *average* admixture of Caucasian genes for American Negroes is between 20% and 30%, there should be enough variance to make it possible to assign large numbers of individuals to at least several categories according to their amount of admixture, and the probable error in classification could be quite definitely specified. A sufficient number of blood groups or other genetic polymorphisms with known frequency distributions in African and Caucasian populations would have to be employed to ensure a high degree of statistical certainty that the categories represented different degrees of genetic racial admixture. A wide range of admixtures probably exists among Negroes living in highly similar environments, so that it should be quite possible in such a study to obtain samples which do not differ across the admixture categories in a number of socioeconomic or other environmental indices. What about skin color? It is polygenetic and is very imperfectly correlated with the amount of Caucasian admixture. Individuals, for example, whose genes are derived in equal (50-50) proportions from African and Caucasian ancestors evince the full range of skin colors from white to black, including all the shades between. This makes it possible statistically to control the effect of skin color; that is, one can compare a number of persons all of whom have the same skin color but different degrees of African/Caucasian admixture, or conversely, the same degree of admixture but different skin colors. (Skin color can be quantified precisely and objectively by means of a photoelectric device which measures reflectance.) The question, then, would be: do the mean IQs (or any other mental ability tests) of the several categories of racial admixture differ significantly and systematically? The genetic equality hypothesis would predict no difference; the genetic inequality hypothesis would predict a difference between the groups.

A further refinement, in order to ensure greater equality of environmental conditions across the admixture categories, *including* prenatal environment, would be to include in the study a large number of half-siblings all related through the mother and reared together. Some half-siblings will inevitably fall into different admixture categories. Do they differ significantly on mental tests

when skin color is controlled? Birth order, maternal age, and other factors would have to be noted, but in large samples these factors would probably tend to be random with respect to racial admixture. One would also want a white control group with no African admixture in order to rule out the remote possibility that the blood groups themselves are causally related to IQ, since they are intended in this study only as genetic markers or indices of racial admixture. Such a study would go further toward answering the question of Negro-white genetic differences in intelligence than the sum total of all the other studies that we now have.

The possibility has been suggested of using genetic linkages for studying the inheritance of intelligence and race differences, but evaluation of its potential merits will have to be decided by geneticists. If the genes for some clearly identifiable physical trait are located on the same chromosome as the genes for some measurable mental ability, we should expect to find a marked correlation in the population between the appearance of the physical characteristic and the mental attribute whose genes share the same chromosome. The physical characteristic would thus serve as an objective genetic marker for the mental trait.

The major difficulty with this approach may be that what we call intelligence is so polygenetic that the relevant genes are carried on most or all of the chromosomes, so that specific linkages could never be established. If intelligence consists of a large number of subabilities, each of which is conditioned independently by a very limited number of genes which are carried on a single chromosome, then it may be possible to study linkages, provided we can reliably measure the subabilities. I have described elsewhere how psychologists might make their measurements of abilities of greater interest and value to researchers in genetics (Jensen 1968b). Briefly, it would consist of the fractionation of mental abilities to the most extreme limits that reliability of measurement will permit, and then seeing if these subabilities show any signs of relatively simple genetic inheritance (such as showing Mendelian ratios) or genetic linkages.

Are there any known linkages between physical and mental characteristics in the normal distribution of intelligence? I do not know of any established examples. We should begin looking for such possible mental linkages with blood groups, biochemical variations, and other physical traits. One set of interesting findings concerns the association between uric acid level in the blood and intellectual achievement. Whether this is an instance of genetic linkage or whether there is a causal connection between uric acid and brain functions is not yet established. Stetten and Hearon (1958) reported a correlation between serum uric acid concentration and scores on the Army intelligence test of 817 inductees. A study of serum urate levels of 51 University of Michigan professors found a positive correlation with drive, achievement, and leadership (Brooks and Mueller 1966), and high school students have been found to show a similar relationship (Kasl, Brooks, and Cobb 1966). It would be interesting to know if these correlations are found within other racial groups and also if there are differences between groups in serum uric acid levels. Every bit of such various kinds of information, if it points consistently in the same direction, reduces to some extent the heredity-environment uncertainty.

There are other promising approaches to this problem through biometrical genetics, but explication of the technical aspects of these methods is clearly beyond the possible scope of the present discussion.

IMPLICATIONS FOR EDUCATION

Since educators have at least officially assumed that race and social class differences in scholastic performance are not associated with any genetic differences in growth rates or patterns of mental abilities but are due entirely to discrimination, prejudice, inequality of educational opportunity, and factors in the child's home environment and peer culture, we have collectively given little if any serious thought to whether we would do anything differently if we knew in fact that all educational differences were not due solely to these environmental factors.

There have been and still are obvious environmental inequities and injustices which have disfavored certain minorities, particularly Negroes, Mexican-Americans, and American Indians. Progress has been made and is continuing to be made to improve these conditions. But there is no doubt still a long way to go, and the drive toward further progress in this direction should be given

top priority in our national effort. Education is one of the chief instruments for approaching this goal. Every child should receive the best education that our current knowledge and technology can provide. This should not imply that we advocate the same methods or the same expectations for all children. There are large individual differences in rates of mental development, in patterns of ability, in drives and interests. These differences exist even among children of the same family. The good parent does his best to make the most of each child's strong points and to help him on his weak points but not make these the crux of success or failure. The school must regard each child, and the differences among children, in much the same way as a good parent should do.

I believe we need to find out the extent to which individual differences, social class differences, and race difference in rates of cognitive development and differential patterns of relative strength and weakness in various types of ability are attributable to genetically conditioned biological growth factors. The answer to this question might imply differences in our approach to improving the education of all children, particularly those we call the disadvantaged, for many of whom school is now a frustrating and unrewarding experience.

Individuals should be treated in terms of their individual characteristics and not in terms of their group membership. This is the way of a democratic society, and educationally it is the only procedure that makes any sense. Individual variations within any large socially defined group are always much greater than the average differences between groups. There is overlap between groups in the distributions of all psychological characteristics that we know anything about. But dealing with children as individuals is not the greatest problem. It is in our concern about the fact that when we do so, we have a differentiated educational program, and children of different socially identifiable groups may not be proportionately represented in different programs. This is the "hang-up" of many persons today and this is where our conceptions of equal opportunity are most likely to go awry and become misconceptions.

Group racial and social class differences are first of all individual differences, but the causes of the *group* differences may not be the same as of the *individual* differences. This is what we must find out, because the prescription of remedies for our educational ills could depend on the answer.

Let me give one quite hypothetical example. We know that among middle-class white children, learning to read by ordinary classroom instruction is related to certain psychological developmental characteristics. Educators call it "readiness." These characteristics of readiness appear at different ages for different kinds of learning, and at any given age there are considerable individual differences among children, even among siblings reared within the same family. These developmental differences, in middle-class white children, are largely conditioned by genetic factors. If we try to begin a child too early in reading instruction, he will experience much greater difficulty than if we waited until we saw more signs of "readiness." Lacking readiness, he may even become so frustrated as to "turn off" on reading, so that he will then have an emotional block toward reading later on when he should have the optimal readiness. The readiness can then not be fully tapped. The child would have been better off had we postponed reading instruction for six months or a year and occupied him during this time with other interesting activities for which he was ready. Chances are he would be a better reader at, say, 10 or 11 years of age for having started a year later, when he could catch on to reading with relative ease and avoid the unnecessary frustration. It is very doubtful in this case that some added "enrichment" to his preschool environment would have made him learn to read much more easily a year earlier. If this is largely a matter of biological maturation, then the time at which a child is taught in terms of his own schedule of development becomes important. If, on the other hand, it is largely a matter of preschool environmental enrichment, then the thing to do is to go to work on the preschool environment so as to make all children equally ready for reading in the first grade. If a child's difficulty is the result of both factors, then a combination of both enrichment and optimal developmental sequencing should be recommended.

There is a danger that some educators' fear of being accused of racial discrimination could become so misguided as to work to the disadvantage of many minority children. Should we deny differential educational treatments to children when such treatment will maximize the benefits they receive from schooling, just because differential treatment might result in disproportionate representation of different racial groups in various programs? I have seen instances where Negro

children were denied special educational facilities commonly given to white children with learning difficulties, simply because school authorities were reluctant to single out *any* Negro children, despite their obvious individual needs, to be treated any differently from the majority of youngsters in the school. There was no hesitation about singling out white children who needed special attention. Many Negro children of normal and superior scholastic potential are consigned to classes in which one-fourth to one-third of their classmates have IQs below 75, which is the usual borderline of educational mental retardation. The majority of these educationally retarded children benefit little or not at all from instruction in the normal classroom, but require special attention in smaller classes that permit a high degree of individualized and small group instruction. Their presence in regular classes creates unusual difficulties for the conscientious teacher and detracts from the optimal educational environment for children of normal ability. Yet there is reluctance to provide special classes for these educationally retarded children if they are Negro or Mexican-American. The classrooms of predominantly minority schools often have 20% to 30% of such children, which handicaps the teacher's efforts on behalf of her other pupils in the normal range of IQ. The more able minority children are thereby disadvantaged in the classroom in ways that are rarely imposed on white children for whom there are more diverse facilities. Differences in rates of mental development and in potentials for various types of learning will not disappear by being ignored. It is up to biologists and psychologists to discover their causes, and it is up to educators to create a diversity of instructional arrangements best suited to the full range of educational differences that we find in our population. Many environmentally caused differences can be minimized or eliminated, given the resources and the will of society. The differences that remain are a challenge for public education. The challenge will be met by making available more ways and means for children to benefit from schooling. This, I am convinced, can come about only through a greater recognition and understanding of the nature of human differences.

NOTE

[1] For example, one need not accept the IQ scale as the most appropriate. If it could be argued and demonstrated that some transformation of the IQ scale produced more orderly and lawful data in studies of heritability, in the degree of normality of the distribution of scores, and in more closely approximating a genetic model, then such a transformation would be justified. It could very well affect the variances of the distributions in different population subgroups. Berkeley geneticist Dr. Jack King, for example, has suggested that if we assume that the factors (genetic and environmental) that affect intelligence do not behave additively but interact multiplicatively (i.e., a factor adds or subtracts a given *percentage* to the total measure rather than a fixed *amount*) a logarithmic transformation of the IQ scale is theoretically justified. In the multiplicative model, the logarithm of the observed measure is normally distributed. The logarithmic transformation in fact makes the IQ distribution more normal (Gaussian) in a number of studies, and it tends to equalize the variances of the Negro and white distributions, although it also has the effect of pulling their means slightly further apart. The proper transformation is 100 (1 + ln IQ/100), which leaves the general population mean at IQ 100. (ln is the natural logarithm.) Past studies of the heritability of intelligence should be re-analyzed using this logarithmic transformation of the IQ scale to see if it gives a closer and more parsimonious fit to a polygenic model.

REFERENCES CITED

Bajema, C. J.

1963 Estimation of the Direction and Intensity of Natural Selection in Relation to Human Intelligence by Means of the Intrinsic Rate of Natural Increase. Eugen. Quart. 10:175-187.

1966 Relation of Fertility to Educational Attainment in a Kalamazoo Public School Population: A Follow-up Study. Eugen. Quart. 13:306-315.

Baratz, S. S., and Joan C. Baratz

1969 Early Childhood Intervention: The Social Science Base of Institutional Racism. Paper presented to the Society for Research in Child Development, Santa Monica, Calif., March.

Bronfenbrenner, U.

1967 The Psychological Costs of Quality and Equality in Education. Child Developm. 38:909-926.

Brooks, G. W., and E. Muller
1966 Serum Urate Concentrations Among University Professors. J. Amer. Med. Assn. 195:415-418.
Burt, C.
1958 The Inheritance of Mental Ability. Amer. Psychol. 13:1-15.
Carter, C. O.
1962 Human Heredity. Baltimore: Penguin Books.
Coleman, J. S., et al.
1966 Equality of Educational Opportunity. US Dept. of Health, Education, and Welfare.
Crow, J. F.
1969 Genetic Theories and Influences: Comments on the Value of Diversity. Harvard Educational Review 39:301-309.
Duncan, O. D., D. L. Featherman, and Beverly Duncan
1968 Socioeconomic Background and Occupational Achievement: Extensions of a Basic Model. Final Report, Project No. 5-0074 (EO-191) US Dept. of Health, Education, and Welfare, Office of Education, Bureau of Research, May.
Higgins, J., S. Reed and E. Reed
1962 Intelligence and Family Size: A Paradox Resolved. Eugen. Quart. 9:84-90.
Jensen, A. R.
1967 Estimation of the Limits of Heritability of Traits by Comparison of Monozygotic and Dizygotic Twins. Proc. Nat. Acad. Sci. 58:149-157.
1968a Social Class, Race, and Genetics: Implications for Education. Amer. Educ. Res. J. 5:1-42.
1968b Another Look at Culture-Fair Tests. *In* Western Regional Conference on Testing Problems, Proceedings for 1968. "Measurement for Educational Planning," pp. 50-104. Berkeley: Educational Testing Service, Western Office.
1969a How Much Can We Boost IQ and Scholastic Achievement? Harvard Educ. Rev. 39:1-123.
1969b Reducing the Heredity-Environment Uncertainty. *In* Environment, Heredity, and Intelligence. Harvard Educ. Rev. Reprint Series No. 2, pp. 209-243.
Kasl, S. V., G. W. Brooks, and S. Cobb
1966 Serum Urate Concentrations in Male High School Students. J. Amer. Med. Assn. 198:713-716.
Kennedy, W. A., V. Van de Riet, and J. C. White
1963 A Normative Sample of Intelligence and Achievement of Negro Elementary School Children in the Southeastern United States. Monogr. Soc. Res. Child Developm. 28(6).
Kuttner, R. E.
1968 Letters to and from the Editor. Perspect. Biol. Med. 11:707-709.
Lesser, G. S., G. Fifer, and D. H. Clark
1965 Mental Abilities of Children from Different Social-Class and Cultural Groups. Monogr. Soc. for Res. in Child Developm. 30(4).
Moynihan, D. P.
1966 Employment, Income, and the Ordeal of the Negro Family. *In* The Negro American, pp. 134-159. T. Parsons and K. B. Clark, eds. Cambridge: Houghton-Mifflin.
Reed, T. E.
1969 Caucasian Genes in American Negroes. Unpublished manuscript, March.
Shuey, Audrey M.
1966 The Testing of Negro Intelligence, Second edition. New York: Social Science Press.
Spuhler, J. N., and G. Lindzey
1967 Racial Differences in Behavior. *In* Behavior-Genetic Analysis. J. Hirsch, ed. New York: McGraw-Hill.
Stanley, J. C., and A. C. Porter
1967 Correlation of Scholastic Aptitude Test Score with College Grades for Negroes Versus Whites. J. Educ. Meas. 4:199-218.
Stetten, D., Jr., and J. Z. Hearon
1959 Intellectual Level Measured by Army Classification Battery and Serum Acid Concentration. Science 129:1737.
Stinchcombe, A. L.
1969 A Critique of Arthur R. Jensen's "How Much Can We Boost IQ and Scholastic Achievement?" Harvard Educ. Rev. 39(3).
Stodolsky, S. S., and G. Lesser
1967 Learning Patterns in the Disadvantaged. Harvard Educ. Rev. 37:546-593.
Tanner, J. M.
1965 The Trend Towards Earlier Physical Maturation. *In* Biological Aspects of Social Problems, pp. 40-66. J. E. Meade and A. S. Parkes, eds. New York: Plenum Press.
1968 Earlier Maturation in Man. Sci. Amer. 218:21-28.

Terman, J. M., and M. Oden
1959 The Gifted Group at Mid-Life. Stanford: Stanford University Press.
Tyler, Leona E.
1965 The Psychology of Human Differences, Third edition. New York: Appleton-Century-Crofts.
Willerman, L., and J. A. Churchill
1967 Intelligence and Birth Weight in Identical Twins. Child Developm. 38:623-629.

This article has been published in:

Brace, C.L., Gamble, G.R., & Bond, J.T. (Eds.), Race and Intelligence. Anthropological Studies Number 8. Washington, D.C.: American Anthropologi Association, 1971.

It also appears in:

Hellmuth, J. (Ed.), Disadvantaged Child, Vol. 3: Compensatory Education: A National Debate. New York: Brunner/Mazel, 1970. Pp. 124-157.

DO SCHOOLS CHEAT MINORITY CHILDREN? *

ARTHUR R. JENSEN *University of California, Berkeley*

Large representative samples of Negro and Mexican-American children from kindergarten to eighth grade in largely *de facto* segregated schools were compared with white children in the same California school district on a comprehensive battery of tests of mental abilities and of scholastic achievement, in addition to personality inventories and indices of socio-economic and cultural disadvantage. It was found that when certain ability and background factors, over which the schools have little or no influence, are statistically controlled, there are no appreciable differences between the scholastic achievements (as measured by the Stanford Achievement Tests) of minority and majority pupils. Moreover, there was no evidence of a 'cumulative deficit' (an increasing gap from lower to higher grade levels between the mean achievements of minority and majority pupils), when the majority-minority differences were measured in standard deviation units. It is concluded that schools in the district under consideration do not cheat minority students in terms of conventional educational criteria. But it might be concluded that minority children are, in fact, cheated if it were shown that their ability patterns require different instructional approaches to optimize their scholastic learning. Marked differences, not only in overall level of ability but also in the *pattern* of abilities, were found between all three of the ethnic groups in this study.

Americans' faith in education is tangibly substantiated in the fact that the American people now invest in educational institutions annually almost as much as all other nations combined. In the past two decades educational spending nationwide has increased fivefold while personal consumption merely doubled. Since World War II school enrolments have increased 86 per cent, while school expenditures (in constant dollars) increased 350 per cent. While employment in private industry increased 38 per cent, it increased 203 per cent in public education. With such an abundant outlay for education, the question naturally arises whether the benefits are equitably distributed to all segments of our population. A keystone of public education is the promise that no child should be denied the opportunity to fulfil his educational potential, regardless of his national, ethnic, or socio-economic background. When substantial inequalities in educational achievement are evident between large segments of the population nominally sharing the same educational system, serious questions are raised, and rightly so. Numerous attempts have been and are being made to find the answers to the inequities in the benefits of education. In California the chief sub-population differences in scholastic attainments involve majority-minority differences, the minorities in this case being Negroes and Mexican-Americans.

The causes of educational inequalities, in terms both of input and output, cannot be discussed very fruitfully in general terms. There are considerable regional and local differences in educational expenditures and facilities and in their distribution within local districts. In assessing the existence and degree of educational inequities, we must get down to specific cases. That is what is intended in this article. We shall

**Educational Research*, 1971, 14, pp. 3–28.

take a rather close look at some of the questions and answers involved in assessing inequalities within a single school system which serves three sub-populations: a majority group, which we shall refer to as Anglos, and two sizeable minorities, Negroes and Mexican-Americans. Before going into the details of this study, however, a few more general points should be reviewed.

School comparisons of academic achievement

The now famous Coleman Report (Coleman, *et al.*, 1966), which surveyed 645,000 pupils in more than 3,000 schools in all regions of the United States, found relatively minor differences in the measured characteristics of schools attended by different racial and ethnic groups, but very great differences in their achievement levels. The Report also argued that when the social background and attitudes of students are held constant, per pupil expenditures, pupil-teacher ratio, school facilities and curricula show very little relation to achievement. The Report concluded '. . . that schools bring little influence to bear on a child's achievement that is independent of his background and general social context' (p. 325). A critical examination of this study by Bowles and Levin (1968) led them to the conclusion that Coleman's methodology could have resulted in an underestimation to some unknown degree of the extent of the relationship between school differences and pupil achievement. They also criticize the conclusion of the Coleman Report that 'There is a small positive effect of school integration on the reading and mathematics achievement of Negro pupils after differences in the socio-economic background of the students are accounted for' (pp. 29-30). Bowles and Levin claim that '. . . the small residual statistical correlation between proportion white in the schools and Negro achievement is likely to be due, at least in part, to the fact that the proportion white in a school is a measure of otherwise inadequately controlled social background of the Negro student. Thus, we find that the conclusion that Negro achievement is positively associated with the proportion of fellow students who are white, once other influences are taken into account, is not supported by the evidence presented in the Report.' Here then is one critique of the Coleman Report which suggests just the opposite of the most popularly held conceptions of what was proved by the Report. Bowles and Levin argue that school effects are probably larger than suggested by the study, and racial composition of the school *per se* is probably a more negligible factor than suggested in the Report's conclusions. A smaller-scale but statistically more thoroughly controlled study by Wilson (1967) found that after controlling for other factors, the racial composition of the school had no significant direct association with Negro achievement, thus supporting the conclusion of Bowles and Levin, at least in the one California school district studied by Wilson.

But probably the most compelling argument for requiring racial balance in schools is not the direct effect of a school's racial composition *per se*, but the fact that it could lead to a greater equalization of school facilities for majority and minority groups so that disadvantaged minorities would not be largely confined to schools with inferior resources. This may be a valid argument in some parts of the country, but one may justifiably question whether it is a cogent factor in California schools.

Consider the following evidence. A rather coarse-grained analysis of the relationship between the proportion of minority enrolment and certain school characteristics in California is made possible by the State Department of Education's recent publication of statistics on several scholastic variables for all school districts in the State. The present analysis, carried out by the writer, is based on only the total of 191 school districts in the ten counties of the greater Bay Area.[1]

The variables on which all school districts were ranked were: *Grade 6 Reading Achievement* (age 11), *Grade 10 Reading* (age 15), *Grade 6 median IQ*, *Grade 10 median IQ*, *Proportion of Minority Enrolment*, *Per Pupil Expenditure*, *Teacher Salary*, *Teacher-Pupil Ratio* (Grades 4-8), *Number of Administrators per 100 Pupils*, and *General Purpose Tax Rate*

[1]Alameda, Contra Costa, Marin, Napa, San Francisco, San Joaquin, San Mateo, Santa Clara, Solano, Sonoma.

Table 1: Correlations (Spearman's ρ) among ten educational variables in 191 California school districts (decimals omitted)

Variable	2	3	4	5	6	7	8	9	10
1. Grade 6 Reading (age 11)	81	94	87	—73	23	21	18	19	—10
2. Grade 10 Reading (age 15)		75	90	—70	08	06	02	—05	—06
3. Grade 6 IQ			85	—67	25	21	17	19	—08
4. Grade 10 IQ				—67	05	05	00	—13	00
5. Minority Enrolment					02	05	08	—10	17
6. Per Pupil Expenditure						35	53	42	47
7. Tax Rate							54	—06	24
8. Teacher Salary								18	45
9. Teacher/Pupil Ratio									01
10. No. Administrators/100									

in the school district. The rank order correlations[a] among these variables for the 191 school districts are shown in Table 1. We see that minority enrolment has quite negligible correlations with all the school facility variables except number of administrators per 100 pupils (Variable 10), and this correlation is positive. On the other hand, there is a strong negative correlation between minority enrolment and the 6th and 10th grade reading and IQ scores. This correlation matrix can be elucidated by factor analysing it, thereby reducing it to three independent components which account for most of the variance (78 per cent). This was accomplished by a varimax rotation of the first three principal components. The rotated factors are shown in Table 2. Factor I is scholastic aptitude (IQ), reading achievement and minority enrolment. Factor II represents the financial resources of the schools, with the highest loading on teacher salary. Factor III is teacher/pupil ratio and that part of per pupil expenditure not associated with Factor II. What this analysis shows most clearly is the absence of any appreciable correlation between the aptitude-achievement variables and the school district's financial outlay. If there were a substantial relationship between the financial resources and the reading achievement of the various school districts, the factors shown in Table 2 could not be so clearly separated. Note also that while minority enrolment has a negative correlation (—·82) with Factor I (IQ-Reading), it has a small positive correlation (+·19) with Factor II (expenditures). The negative correlation (—·09) between minority enrolment and Factor III indicates a slight disadvantage to districts with a high proportion of minorities in terms of average class size. Overall, these data suggest that there is no appreciable relationship between these particular school resources and minority enrolment, and if anything the correlation is in just the opposite direction to the popular belief that educational facilities are relatively inadequate in districts with a higher percentage of minority students.

Since this analysis is based on data in which the smallest unit for analysis is the school district, it permits no inference concerning the allocation of educational resources to the various schools, which probably differ in minority enrolments, within the districts. A similar analysis could be performed within a district, using the individual schools as the unit of analysis, but different

Table 2: Rotated factor loadings for ten educational variables in 191 California school districts

Variables	Factors I	II	III
1. Grade 6 Reading (age 11)	·95	·12	·15
2. Grade 10 Reading (age 15)	·92	·09	—·08
3. Grade 6 IQ	·92	·13	·17
4. Grade 10 IQ	·95	·06	—·17
5. Minority Enrolment	—·82	·19	—·09
6. Per Pupil Expenditure	·10	·67	·55
7. Tax Rate	·11	·75	—·15
8. Teacher Salary	·05	·88	·17
9. Teacher/Pupil Ratio	·08	·01	·96
10. No. of Administrators	—·13	·71	·01
Per Cent of Variance	42·4	22·6	13·6

[a]A smaller rank order (e.g. 1) indicates: high reading scores, high median IQ, high proportion of minorities, high expenditure per child, high teacher salaries, high tax rate, high teacher/pupil ratio (i.e., smaller classes), and a larger number of administrators per 100 pupils.

indices of a school's resources would have to be used, since there would be relatively little variance on such variables as teacher salary and per pupil expenditure *within* any given school district. More fine-grained indices of the school's specific educational facilities should be included. In any case, the first and most obvious step in assessing the equality of educational facilities is to make a direct examination of the facilities, per pupil expenditures, etc. The recreational, hygienic, safety, and aesthetic aspects of the school plant should be considered no less than those facilities deemed to have more direct educational consequences, such as pupil/teacher ratio and special services.

The misuse of national and state norms

School boards, the public, and the press commonly misuse the published and state norms on standardized achievement tests. Schools and districts are compared against 'norms,' which are intended to represent national or state averages, as if achieving a close approximation to the norms, if not exceeding them, should be the primary goal of every school system. Deviation from the norm, above or below, is commonly regarded as a credit or a discredit to the particular school system. The fallacy in this, of course, is the fact that the average level of scholastic achievement in a community is highly predictable from a number of the community's characteristics over which the local schools have no control whatsoever. Thorndike (1951), for example, correlated average IQ and an average scholastic achievement index (based on half a million children) with 24 census variables for a wide range of communities, large and small, urban and rural. Eleven of the correlations were significant at the one per cent level. Census variables with the highest correlation with IQ and achievement were educational level of the adult population (·43), home ownership (·39), quality and cost of housing (·33), proportion of native-born whites (·28), rate of female employment (·26), and proportion of professional workers (·28). In a multiple correlation these census variables predicted IQ and achievement between ·55 and ·60. Essentially the same picture is revealed in many other similar studies (Wiseman, 1964, Chapter IV). A school's or district's deviation from the mean achievement predicted from a multiple regression equation based on a host of community characteristics would, therefore, make much more sense than a mere comparison of the school's average with national or state norms.

Majority-minority comparisons within a school district

Even when a school district has equalized the educational facilities in all of its schools in terms of physical plant amenities, teacher salaries and qualifications, per pupil expenditures, teacher/pupil ratios, special services, curriculum, and the like, the question may still be asked whether majority-minority differences in scholastic achievement are a product of more subtle and less tangible factors operating in the school situation. We have in mind, for example, such factors as racial and socio-economic composition of the school, and different teacher attitudes and expectancies in relation to majority and minority pupils. Is there any way we can assess the degree to which schools afford unequal educational advantages to majority and minority pupils over and above what can easily be reckoned in terms of pupil expenditures and the like?

I have tried to answer this question as best as I believe it can be answered with the psychometric and statistical methodology now available and with the rather modest resources within the financial means of most school systems. Although it would be impossible to present all the technical details and results of this study within the limits of this paper, it is possible to indicate some of the methods and the most relevant results they have yielded.

The study was conducted in 1970 in a fairly large (35 schools) elementary school district of California. This school district was ideal for this kind of study for four main reasons: (1) the district's school population has substantial proportions of Negro (13 per cent) and Mexican-American (20 per cent) students; (2) the majority (Anglo) population is very close to state and national norms for Anglos in IQ, for both mean and standard deviation, and the same is true for the two minority groups in relation to norms for their respective populations in the US; (3) the schools are largely *de facto* segregated due to rather

widely spaced residential clustering of the three ethnic groups, and (4) the district had made a thorough effort to provide equal educational facilities in all of its schools, if anything favouring those schools with the largest minority enrolments to whom additional federal and state funds were allocated for special compensatory programmes.

Large representative samples totalling 28 per cent of the school population from kindergarten to the eighth grade (age 13) were selected for study. A total of 6,619 children were tested; more or less equal numbers were tested at each grade. The three main ethnic classifications were Anglo ($N = 2{,}453$), Mexican-American ($N = 2{,}263$), and Negro ($N = 1{,}853$). Approximately half the sample (selected randomly with the classroom as the unit of selection) were tested by a small staff of specially trained testers, and half were tested by their regular classroom teachers. Because of the large sample sizes the tester and teacher results often differ significantly but do not differ appreciably or systematically except that the results of teacher-administered tests consistently have somewhat greater variance and lower reliability which would tend to attenuate intercorrelations among measures and lessen the statistical significance of group differences. Parallel analyses for testers and teachers were run on all the data, which were combined when there were no significant or systematic differences between the two forms of testing. For the sake of simplicity in the present summary only the tester results are reported here when the two sets of data were not combined.

Rationale of the study

In terms of this study one can think of the educational process as being analogous to an industrial production process in which raw materials ('input') are converted to a specified product ('output'). The output will be a function both of the input and of the effectiveness of the process by means of which the input is converted into output. In the case of schooling, the input is what the child brings with him to school by way of his abilities, attitudes, prior learning, cultural background, and personality characteristics relevant to learning in the classroom. The school itself has relatively little, if any, control over these input variables. The school, however, can have considerable influence on one variable—prior learning—for children who are already somewhere along the educational path, and if the school's instructional programme is deficient for some children, the deficiencies in prior learning in earlier grades should show up increasingly in later grades as a cumulating deficit in scholastic achievement.

Whatever else one may say about it, schooling is essentially a process whereby children are helped to acquire certain skills, which are the output of the system. The effectiveness of the process can be judged, among other ways, in terms of the relationship between input and output. Meaningful comparisons cannot be made between the output (scholastic achievement) of different pupils, classes, schools, or school districts without reference to the input variables. The main purpose of the present study is the comparison of the outputs, i.e., educational achievements, of three categories of pupils—Anglo, Negro, and Mexican-American—when these groups are statistically equated on the input variables. In this way we can make some judgement concerning the relative efficiency of the educational process for each of the three groups. The adequacy of the statistical equating of the groups in terms of input depends upon a judicious selection of instruments for measuring the input variables. The chief aims in selecting the input control variables are (1) to represent the domain of educationally relevant abilities, personality, and home background factors as broadly as feasible, and (2) to include only those ability and background variables which are not explicitly taught by the schools or are not under direct control of the schools. That is to say, they should represent the raw materials that the schools have to work with. The output, on the other hand, should represent objective measures of those skills which it is the school's specific purpose to teach. These are best measured by standardized tests of scholastic achievement.

The input variables can be classified into three categories: (1) ability or general aptitude tests, (2) motivation, personality, and school-related attitudes, and (3) environmental background variables reflecting socio-economic status, parental education, and general cultural advantages.

Input variables

Ability Tests

Lorge-Thorndike Intelligence Test. This is a nationally standardized group-administered test of general intelligence. In the normative sample, which was intended to be representative of the nation's school population, the test has a mean IQ of 100 and a standard deviation of 16. It is generally acknowledged to be one of the best paper-and-pencil tests of general intelligence.

The Manual of the Lorge-Thorndike Test states that the test was designed to measure reasoning ability. It does not test proficiency in specific skills taught in school, although the verbal tests, from grade 4 (age nine) and above, depend upon reading ability. The reading level required, however, is intentionally kept considerably below the level of reasoning required for correctly answering the test questions. Thus the test is essentially a test of reasoning and not of reading ability, which is to say that it should have more of its variance in common with non-verbal tests of reasoning ability than with tests of reading *per se*.

The tests for grades K–3 do not depend at all upon reading ability but make use exclusively of pictorial items. The tests for grades 4-8 consist of two parts, *Verbal* (V) and *Non-verbal* (NV). They are scored separately and the raw score on each is converted to an IQ, with a normative mean of 100 and SD of 16. The chief advantage of keeping the two scores separate is that the Non-verbal IQ does not overestimate or under-estimate the child's general level of intellectual ability because of specific skills or disabilities in reading. The Non-verbal IQ, however, correlates almost as highly with a test of reading comprehension as does the Verbal IQ, because all three tests depend primarily upon reasoning ability and not upon reading *per se*. For example, in the fourth grade sample, the correlation between the Lorge-Thorndike Verbal and Non-Verbal IQs is ·70. The correlation between Verbal IQ and the Paragraph Meaning Sub-test of the Standard Achievement Test is ·52. The correlation between the Non-verbal IQ and Paragraph Meaning is ·47. Now we can ask: what is the correlation of Verbal IQ and Paragraph Meaning when the effects of Non-verbal IQ are partialled out, that is, are held constant? The partial correlation between Verbal IQ and Paragraph Meaning (holding Non-verbal IQ constant) is only ·29.

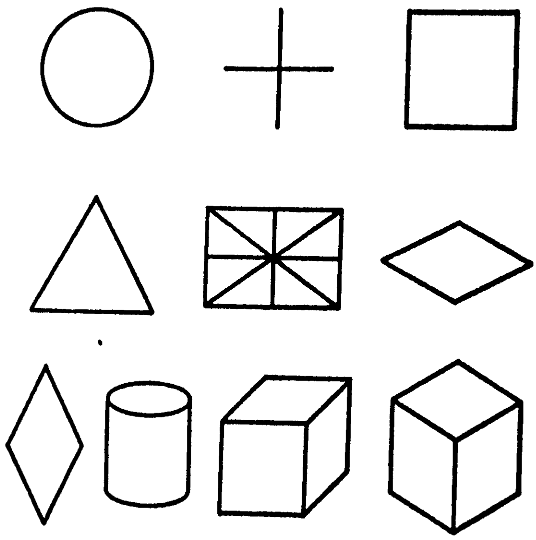
The following forms of the Lorge-Thorndike Intelligence Tests were used:

Level 1, Form B		Grades K–1
Level 2, Form B		Grades 2–3
Level 3, Form B	Verbal and Non-verbal	Grades 4–6
Level 4, Form B	Verbal and Non-verbal	Grades 7–8

Figure Copying Test. The Figure Copying Test was given in grades K–6. Beyond grade 6 (age 11) too large a proportion of children obtain the maximum possible score (30) for the test to be useful in making group comparisons. In fact, by grades 5 and 6 group differences are very probably underestimated by this test, since a larger proportion of the higher-scoring group will obtain the maximum score and this 'ceiling' effect will prevent the group's full range of ability from being represented. The ceiling effect consequently spuriously depresses the group's mean and reduces the variance (or standard deviation). Nevertheless, this test is extremely valuable for group comparisons because it is one of the least culture-loaded tests available and successful performance on the test is known to be significantly related to readiness for the scholastic tasks of the primary grades, especially reading readiness.

The Figure Copying Test was developed at the Gesell Institute of Child Study at Yale University as a means of measuring developmental readiness for the traditional school learning tasks of the primary grades. The test consists of the ten geometric forms shown in Figure 1, arranged in order of difficulty, which the child must simply copy, each on a separate sheet of paper. The test involves no memory factor, since the figure to be copied is before the child at all times. The test is administered without time limit, although most children finish in 10 to 15 minutes. The test is best regarded as a developmental scale of mental ability. It correlates substantially with other IQ tests, but it is considerably less culture-loaded than most usual IQ tests. It is primarily a measure of general cognitive development and not just of perceptual-motor ability. Children

Figure 1: The ten simple geometric forms used in the Figure Copying Test



In the actual test booklet each figure is present singly in the top half of a 5½" × 8½" sheet. The circle is 1⅝" in diameter.

taking the test are urged to attempt to copy every figure.

Each of the ten figures is scored on a three-point scale going from one (low) to three (high). (A score of zero is given in the rare instance when no attempt has been made to copy a particular figure.) A score of one is given if an attempt is made but the child's drawing completely fails to resemble the model. A score of two is given if there is fair resemblance to the model—the figure need not be perfect but it must be easily recognizable as the model which the child has attempted to copy. A score of three is given for an attempt which duplicates the figure in all its essential characteristics—this is an essentially adult level of performance. Since there are ten figures in all, the possible range of scores goes from 10 to 30 (or 0 to 30 if zeros are counted, but this is rare, since virtually all subjects attempt all ten figures).

The high level of motivation maintained by this test is indicated by the fact that the minimum score obtained in each group at each grade level increases systematically with grade level. This suggests that all children were making an attempt to perform in accordance with the instructions. Another indication that can be seen from the test booklets is that virtually 100 per cent of the children in every ethnic group at every grade level attempted to copy every figure. The attempts, even when unsuccessful, usually show considerable effort, as indicated by re-drawing the figure, erasures and drawing over the figure repeatedly in order to improve its likeness to the model. It is also noteworthy about this test that normal children are generally not successful in drawing figures beyond their mental age level and that special instructions and coaching on the drawing of these figures hardly improves the child's performance. This test, in other words, is not very susceptible to training, but measures some fundamental aspects of mental development. The diagnostic significance of this test has been explicated extensively in *School Readiness* (Harper and Row, 1967, pp. 63-129) by Drs. Frances L. Ilg and Louise Bates Ames of the Gesell Institute of Child Development at Yale University.

Raven's Progressive Matrices. This non-verbal reasoning test, devised in England, is intended to be a pure measure of *g*, the general factor common to all intelligence tests. It is a highly reliable measure of reasoning ability, quite free of the influence of special abilities, such as verbal or numerical facility. It is probably the most culture-free test of general intelligence yet devised by psychologists. The test mainly gets at the ability to grasp relationships; it does not depend upon specific acquired information as do tests of vocabulary, general information, etc. The test, which is group-administered, begins with problems that are so easy that all children by third grade can catch on and solve the problems even without instructions.

Two forms of the test were used. The Coloured Progressive Matrices, which is the children's form, was used in grades 3 to 6. This test is appropriate even for kindergarten children, but to ensure that all children tested could go through the first few problems without difficulty, giving them a chance to catch on easily and experience success in the early part of the test, we used this test only from the third grade and above. The Coloured Matrices consist of 36 matrix problems which are administered without time limit. Children are encouraged to attempt all problems. There is no penalty for guessing.

The Standard Progressive Matrices were used in grades 7 and 8. These begin as easily as the coloured matrices but advance in difficulty more rapidly and go up to a level appropriate for average adults. There are 60 matrix problems in all, and the subjects are encouraged to attempt all of them, without penalty for guessing.

Listening-Attention Test. In the Listening-Attention Test the child is presented with an answer sheet containing 100 pairs of digits in sets of 10. The child listens to a tape recording which speaks one digit every two seconds. The child is required to put an X over the one digit in each pair which has been heard on the tape recorder. The purpose of this test is to determine the extent to which the child is able to pay attention to numbers spoken on a tape recorder, to keep his place in the test, and to make the appropriate responses to what he hears from moment to moment. Low scores on this test indicate that the subject is not yet ready to take the Memory for Numbers test which immediately follows it. High scores on the Listening-Attention Test indicate that the subject has the prerequisite skills for taking the digit span (Memory for Numbers) test. The Listening-Attention Test thus is intended as a means for detecting students who, for whatever reason, are unable to hear and to respond to numbers read over a tape recorder. The test itself makes no demands on the child's memory, but on on his ability for listening, paying attention, and responding appropriately—all prerequisites for the digit memory test that follows.

It has been found in previous studies using the Listening-Attention Test that the vast majority of subjects from grade 2 and above obtain perfect scores; the median score is 100, and the lower quartile rarely goes below 95. This means that nearly all subjects have the prerequisite skills for the Memory for Numbers test to yield a valid measure of the subjects' short-term memory ability.

Memory for Numbers Test. The Memory for Numbers test is a measure of digit span, or more generally, short-term memory. It consists of three parts. Each part consists of six series of digits going from four digits in a series up to nine digits in a series. The digit series are presented on a tape recording on which the digits are spoken clearly by a male voice at the rate of precisely one digit per second. The subjects write down as many digits as they can recall at the conclusion of each series, which is signalled by a 'bong'. Each part of the test is preceded by a short practice test of three-digit series in order to permit the tester to determine whether the child has understood the instructions, etc. The practice test also serves to familiarise the subject with the procedure of each of the sub-tests. The first sub-test is labelled Immediate Recall (I). Here the subject is instructed to recall the series *immediately* after the last digit has been spoken on the tape recorder. The second sub-test consists of Delayed Recall (D). Here the subject is instructed not to write down his response until after ten seconds have elapsed after the last digit has been spoken. The ten-second interval is marked by audible clicks of a metronome and is terminated by the sound of a 'bong' which signals the child to write his response. The Delayed Recall condition

invariably results in some retention loss. The third sub-test is the repeated series test, in which the digit series is repeated three times prior to recall; the subject then recalls the series immediately after the last digit in the series has been presented. Again, recall is signalled by a 'bong'. Each repetition of the series is separated by a tone with a duration of one second. The repeated series almost invariably results in greater recall than the single series. This test is very culture-fair for children in second grade and beyond and who know their numerals and are capable of listening and paying attention, as indicated by the Listening-Attention Test. The maximum score on any one of the sub-tests is 39, that is the sum of the digit series from four through nine.

Motivational and Personality Tests

Speed and Persistence Test (Making X's). The Making X's Test is intended as an assessment of test-making motivation. It gives an indication of the subject's willingness to comply with instructions in a group testing situation and to mobilize effort in following those instructions for a brief period of time. The test involves no intellectual component, although for young children it probably involves some perceptual-motor skills component, as reflected by increasing mean scores as a function of age between grades 1 to 5. The wide range of individual differences among children at any one grade level would seem to reflect mainly general motivation and test-making attitudes in a group situation. The test also serves partly as an index of classroom morale, and it can be entered as a moderator variable into correlational analyses with other ability and achievement tests. Children who do very poorly on this test, it can be suspected, are likely not to put out their maximum effort on ability tests given in a group situation and therefore their scores are not likely to reflect their 'true' level of ability.

The Making X's Test consists of two parts. On Part I the subject is asked simply to make X's in a series of squares for a period of 90 seconds. In this part the instructions say nothing about speed. They merely instruct the child to make X's. The maximum possible score on Part I is 150, since there are 150 squares provided in which the child can make X's. After a two-minute rest period the child turns the page of the test booklet to Part II. Here the child is instructed to show how much better he can perform than he did on Part I and to work as rapidly as possible. The child is again given 90 seconds to make as many X's as he can in the 150 boxes provided. The gain in score from Part I to Part II reflects both a practice effect and an increase in motivation or effort as a result of the motivating instructions, i.e. instructions to work as rapidly as possible.

Ethnic and social-class group differences on this test are generally smaller than on any other test, with the exception of the Listening-Attention Test (on which there are almost no group *or* individual differences).

Eysenck Personality Inventory-Junior. The EPI-Junior is the children's form of the EPI for adults. It is a questionnaire designed to measure the two factors of personality which have been found to account for most of the variance in the personality domain—Extraversion and Neuroticism. The Extraversion (E) scale represents the continuum of social extraversion-introversion. High scores reflect sociability, outgoingness and carefreeness. The Neuroticism (N) scale reflects emotional instability, anxiety proneness, and the tendency to develop neurotic symptoms under stress. The Lie (L) scale is merely a validity detector consisting of a number of items which are very rarely answered in the keyed direction by the vast majority of subjects. A high score on L indicates that the subject is 'faking good' or is answering the questionnaire items more or less at random, either intentionally or as a result of insufficient comprehension of the items. Naivity is also reflected in elevated L scores, and it is probably mainly this factor which causes a decrease in L scores as children mature.

The EPI scales were included in the present study as a control variable because previous studies had shown the E and N scales to predict a small but significant part of the variance in scholastic performance. Because of the reading level required by the EPI, it was not given below the fourth grade.

Student Self-Report. This 21-item self-report inventory was composed mainly of items in the self concept inventory used by James Coleman in his study, *Equality of Educational Opportunity.* It

reveals the student's attitudes towards school, towards himself as a student, and other attitudes affecting motivation and self-esteem. The questionnaire was administered by the classroom teachers to grades 4 to 8. Because of the reading level required, it was not administered below grade 4.

Background Information

The Home Index. This is a 24-item questionnaire about the home environment, devised by Harrison Gough (1949). It is a sensitive composite index of the socio-economic level of the child's family. Factor analysis of past data by Gough has shown that the 24 items fall into four categories, each of which can be scored as a separate scale. Part I (Items 6, 7, 8, 9, 10, 15, 16, 23) reflects primarily the educational level of the parents. Part II (Items 1, 2, 3, 4, 5, 13, 20, 24) reflects material possessions in the home. Part III (Items 17, 18, 21, 22) reflects degree of parental participation in middle or upper-middle class social and civic activities. Part IV (Items 11 and 19) relates to formal exposure to music and other arts.

Output Variables: Scholastic Achievement

Stanford Achievement Tests. Scholastic achievement was assessed by means of the so-called 'partial battery' of the Stanford Achievement Tests, consisting of the following sub-tests: Word Meaning, Paragraph Meaning, Spelling, Word Study Skills, Language (grammar), Arithmetic Computation, Arithmetic Concepts, and Arithmetic Applications. The Stanford Achievement battery was administered to grades 1 to 8.

Distinction between aptitude and achievement

Can we justify the separation of our tests into two categories, ability or aptitude tests versus scholastic achievement tests, and then regard the former as *input* and the latter as *output*? Do not intelligence or aptitude tests also measure learning or achievement? The answer to this question is far from simple, but I believe there are at least six kinds of evidence which justify a psychological distinction between intelligence tests and achievement tests.

(1) *Breadth of learning sampled.* The most obvious difference between tests of intelligence and of achievement is the breadth of the domains sampled by the tests. Achievement tests sample very narrowly from the most specifically taught skills in the traditional curriculum, emphasizing particularly the three R's. Achievement test items are samples of the particular skills that children are specifically taught in school. Since these skills are quite explicitly defined and the criteria of their attainment are fairly clear to teachers and parents, children can be taught and can be given practice on these skills to shape their performance up to the desired criterion. Because of the circumscribed nature of many of the basic scholastic skills, the pupil's specific weaknesses can be identified and remedied. The skills or learning sampled by an intelligence test, on the other hand, represent achievements of a much broader nature. Intelligence test items are sampled from such a very wide range of potential experiences that the idea of teaching intelligence, as compared with teaching, say, reading or arithmetic, is practically nonsensical. Even direct coaching and practice on a particular intelligence test raises individual's scores on the average by only five to ten points; and some tests, especially those referred to as 'culture fair', seem to be hardly amenable to the effects of coaching and practice. The average five year old, for example, can copy a circle or a square without any trouble, but try to teach him to copy a diamond and see how far he gets! Wait until he is seven years old and he will have no trouble copying the diamond without any need for instruction. Even vocabulary is very unsusceptible to enlargement by direct practice aimed at increasing vocabulary. This is part of the reason why vocabulary tests are regarded as such good measures of general intelligence and always have a high *g* loading in factor analyses of various types of intelligence tests. The items in a vocabulary test are sampled from such an enormously large pool of potential items that the number that can be acquired by specific study and practice is only a small proportion of the total, so that few if any are likely to appear in any given vocabulary test. Furthermore, persons seem to retain only those words which fill some conceptual 'slot' or need in their own mental structures. A new word encountered for the first time which fills such a conceptual 'slot' is picked up and retained

seemingly without conscious effort, and will 'pop' into mind again when the conceptual need for it arises, even though in the meantime the word may not have been encountered for many months or even years. If there is no conceptual slot needing to be filled, that is to say, no meaning for the individual which the word serves to symbolize, it is very difficult to make the definition of the word stick in the individual's memory, and even after repeated drill, it will quickly fade beyond retrieval, as when a student memorizes a long list of foreign words in order to pass his foreign language exam. for the PhD. Since intelligence tests assess the learning that occurs in the total life experiences of the individual, they are more general and more valid measures of his learning potential than are scholastic achievement tests. It should come as no surprise that there is a substantial correlation between the two classes of tests, since both measure learning or achievement, one in a broad sphere, the other in a much narrower sphere. In a culturally more or less homogeneous population the broader-based measure called 'intelligence' is more generally representative of the individual's learning capacities and is more stable over time than the more specific acquisitions of knowledge and skill classed as scholastic achievement.

(2) *Equivalence of Diverse Tests.* One of the most impressive characteristics of intelligence tests is the great diversity of means by which essentially the same ability (or abilities) can be measured. Tests having very diverse forms, such as vocabulary, block designs, matrices, number series, 'odd-man out', figure copying, verbal analogies, and other kinds of problems can all serve as intelligence tests yielding more or less equivalent results because of their high intercorrelations. All these types of tests have high loadings on the *g* factor, which, as Wechsler (1958, p. 121) has said, '. . . involves broad mental organization; it is independent of the modality or contextual structure from which it is elicited; *g* cannot be exclusively identified with any single intellectual ability and for this reason cannot be described in concrete operational terms.' We can accurately define *g* only in terms of certain mathematical operations; in Wechsler's words '*g* is a measure of a collective communality which necessarily emerges from the intercorrelation of any broad sample of mental abilities' (p. 123).

Assessment of scholastic achievement, on the other hand, depends upon tests of narrowly specific acquired skills—reading, spelling, arithmetic operations, and the like. The forms by means of which one can test any one of these scholastic skills are very limited indeed. This is not to say that there is not a general factor common to all tests of scholastic achievement, but this general factor common to all the tests seems to be quite indistinguishable from the *g* factor of intelligence tests. Achievement tests, however, usually do not have as high *g* loadings as intelligence tests but have higher loadings on group factors such as verbal and numerical ability factors and they also contain more task-specific variance. It is always possible to make achievement tests correlate more highly with intelligence tests by requiring students to reason, to use data provided, and to apply their factual knowledge to the solution of new problems. More than just the mastery of factual information, intelligence is the ability to apply this information in new and different ways. With increasing grade level, achievement tests have more and more variance in common with tests of *g*. For example, once the basic skills in reading have been acquired, reading achievement tests must increasingly measure the student's comprehension of more and more complex selections rather than the simpler processes of word recognition, decoding, etc. And thus at higher grades, tests of reading comprehension, for those children who have already mastered the basic skills, become more or less indistinguishable in factorial composition from the so-called tests of verbal intelligence. Similarly, tests of mechanical arithmetic (arithmetic computation) have less correlation with *g* than tests of arithmetic thought problems, such as the Arithmetic Concepts and Arithmetic Applications sub-tests of the Stanford Achievement battery. Accordingly, most indices of scholastic performance increasingly reflect general intelligence as children progress in school. We found in our study, for example, that up to grade 6, verbal and non-verbal intelligence tests could be factorially separated, with the scholastic achievement tests lining up on the same factor with verbal intelligence. But beyond grade 6 both the verbal and

non-verbal tests, along with all the scholastic achievement tests, amalgamated into a single large general factor which no form of factor rotation could separate into smaller components distinguishable as verbal intelligence versus non-verbal intelligence versus scholastic achievement. By grades 7 and 8 the Lorge-Thorndike Non-verbal IQ and Raven's Progressive Matrices are hardly distinguishable in their factor composition from the tests of scholastic achievement. At the same time it is important to recognize that the Lorge-Thorndike Non-verbal IQ and Raven's Matrices are not measuring scholastic attainment *per se*, as demonstrated by the fact that totally illiterate and unschooled persons can obtain high scores on these tests. Burt (1961), for example, reported the case of separated identical twins with widely differing educational attainments (elementary school education versus a university degree), who differed by only one IQ point on the Progressive Matrices (127 versus 128).

(3) *Heritability of Intelligence and Scholastic Achievement*. Another characteristic which distinguishes intelligence tests from achievement tests is the difference between the heritability values generally found for intelligence and achievement measures. Heritability is a technical term in quantitative genetics referring to the proportion of test score variance (or any phenotypic variance) attributable to genetic factors. Determinations of the heritability of intelligence test scores range from about ·60 to ·90, with average values around ·70 to ·80 (Jensen, 1969a). This means that some 70 to 80 per cent of the variance in IQs in the European and North American Caucasian population in which these studies have been made is attributable to genetic variance, and only 20 to 30 per cent is attributable to non-genetic or environmental variability. The best evidence now available shows a somewhat different picture for measures of scholastic achievement, which on the average have much lower heritability. A review of all twin studies in which heritability was determined by the same methods for intelligence tests and for achievement tests shows an average heritability of ·80 for the former and of only ·40 for the latter (Jensen, 1967). It is likely that scholastic measures increase in heritability with increasing grade level and that the simpler skills such as reading, spelling, and mechanical arithmetic have lower heritability than the more complex processes such as reading comprehension and arithmetic applications. The reason is quite easy to understand. Simple circumscribed skills can be more easily taught, drilled, and assessed and the degree of their mastery for any individual will be largely a function of the amount of time he spends in being taught and in practising the skill. Thus children with quite different learning abilities can be shaped up to perform more or less equally in these elemental skills. If Johnny has trouble with his reading or arithmetic or spelling his parents may give him extra tutoring so that he can more nearly approximate the performance of his brighter brother. Siblings in the same family differ considerably less in scholastic achievement than in intelligence. Conversely, identical twins reared apart differ much more in scholastic achievement than in intelligence. From these facts we conclude that environmental factors make a larger contribution to individual differences in achievement than in intelligence as measured by standard tests.

(4) *Maturational Aspects of Intelligence*. An important characteristic of the best intelligence test items is that they clearly fall along an age scale. Items are thus 'naturally' ordered in difficulty. The Figure Copying Test (see Figure 1) is a good example. Ability to succeed on a more difficult item in the age scale is not functionally dependent upon success on previous items in the sense that the easier item is a prerequisite component of the more difficult item. By contrast, skill in short division is a component of skill in long division. The age differential for some tasks such as figure copying and the Piagetian conservation tests is so marked as to suggest that they depend upon the sequential maturation of hierarchical neural processes (Jensen, 1970b). Teaching of the skills before the necessary maturation has occurred is often practically impossible, but after the child has reached a certain age successful performance of the skill occurs without any specific training or practice. The items in scholastic achievement tests do not show this characteristic. For successful performance, the subject must have received explicit instruction in the specific subject matter of the test. The teachability of scholastic subjects is much more obvious

than of the kinds of materials that constitute most intelligence tests and especially non-verbal tests.

Cumulative Deficit and the Progressive Achievement Gap

The concept of 'cumulative deficit' is fundamental in the assessment of majority-minority differences in educational progress. Cumulative deficit is actually an hypothetical concept intended to explain an observable phenomenon which can be called the 'progressive achievement gap' or PAG for short. When two groups show an increasing divergence between their mean scores on tests, there is potential evidence of a PAG. The notion of cumulative deficit attributes the increasing difference between the groups' means to the cumulative effects of scholastic learning such that deficiencies at earlier stages make for greater deficiencies at later stages. If Johnny fails to master addition by the second grade he will be worse off in multiplication in the third grade, and still worse off in division in the fourth grade, and so on. Thus the progressive achievement gap between Johnny and those children who adequately learn each prerequisite for the next educational step is seen as a cumulative deficit. There may be other reasons as well for the PAG, such as differential rates of mental maturation, the changing factorial composition of scholastic tasks which means that somewhat different mental abilities are called for at different ages, disillusionment and waning motivation for school work, and so on. Therefore I prefer the term 'progressive achievement gap' because it refers to an observable effect and is neutral with respect to its causes.

Absolute and Relative PAG. When the achievement gap is measured in raw score units or in grade scale or age scale units, it is called *absolute*. For example, we read in the Coleman Report (1966, p. 273) that in the metropolitan areas of the northwest region of the US '... the lag of Negro scores (in Verbal ability) in terms of years behind grade level is progressively greater. At grade 6, the average Negro is approximately 1 1/2 years behind the average white. At grade 9, he is approximately 2 1/4 years behind that of the average white. At grade 12, he is approximately 3 1/4 years behind the average white.'

When the achievement difference between groups is expressed in standard deviation units, it is called *relative*. That is to say, the difference is relative to the variation within the criterion group. The Coleman Report, referring to the findings quoted above, goes on to state: 'A similar result holds for Negroes in all regions, despite the constant difference in number of standard deviations.' Although the absolute white-Negro difference increases with grade in school, the relative difference does not. The Coleman Report states: 'Thus in one sense it is meaningful to say the Negroes in the metropolitan North-east are the same distance below the whites at these three grades—that is, relative to the dispersion of the whites themselves.' The Report illustrates this in pointing out that at grade 6 about 15 per cent of whites are one standard deviation, or 1 1/2 years, behind the white average; at grade 12, 15 per cent of the whites are one standard deviation, of three and a quarter years behind the white average.

It is of course the absolute progressive achievement gap which is observed by teachers and parents, and it becomes increasingly obvious at each higher grade level. But statistically the proper basis for comparing the achievement differences between various sub-groups of the school population is in terms of the relative difference, that is, in standard deviation units, called sigma (σ) units for short.

Except in the southern regions of the US, the Coleman study found a more or less constant difference of approximately one sigma (based on whites in the metropolitan north-east) between whites and Negroes in Verbal Ability, Reading Comprehension, and Maths Achievement. In other words, there was no progressive achievement gap in regions outside the south. In the southern regions, there is evidence for a PAG from grade 6 to 12 when the sigma unit is based on the metropolitan north-east. For example, in the non-metropolitan south, the mean Negro-white differences (Verbal Ability) in sigma units are 1·5, 1·7, and 1·9 for grades 6, 9, and 12, respectively. The corresponding number of grade levels that the southern Negroes lag behind at grades 6, 9, and 12 are 2·3, 3·9, and 5·2 (Coleman, 1966, p. 274). The causes of this progressive achievement gap in the south are not definitely known. Contributing factors could be an actual cumulative deficit in educational skills,

true sub-population differences in the developmental growth rates of the mental abilities relevant to school learning, and selective migration of families of abler students out of the rural south, causing an increasing cumulation of poor students in the higher grades.

Cross-Sectional versus Longitudinal PAG. Selective migration, student turnover related to adult employment trends, and other factors contributing to changes in the characteristics of the school population may produce a spurious PAG when this is measured by comparisons between grade levels at a single cross-section in time. The Coleman Report's grade comparisons are cross-sectional. But where there is no reason to suspect systematic regional population changes, cross-sectional data should yield approximately the same picture as longitudinal data, which are obtained by repeated testing of the same children at different grades. Longitudinal data provide the least questionable basis for measuring the PAG. Cross-sectional achievement data can be made less questionable if there are also socio-economic ratings on the groups being compared. The lack of any grade-to-grade decrement on the socio-economic index adds weight to the conclusion that the PAG is not an artifact of the population's characteristics differing across grade levels. (This type of control was used in the present study reported in the following section.)

Another way of looking at the PAG is in terms of the percentage of variance in individual achievement scores accounted for by the mean achievement level of schools or districts. If there is an achievement decrement for, say, a minority group across grade levels, and if the decrement is a result of school influences, then we should expect an increasing correlation between individual students' achievement scores and the school averages. In the data of the Coleman Report, this correlation (expressed as the percentage of variance in individual scores accounted for by the school average) for 'verbal achievement' does not change appreciably from the beginning of the first school year up to the 12th grade. The school average for verbal achievement is as highly correlated with individual verbal achievement at the beginning of grade 1 as at grade 12. If the schools themselves contributed to the deficit, one would expect an increasing percentage of the total individual variance to be accounted for by the school average with increasing grade level. But no evidence was found that this state of affairs exists. The percentage of total variance in individual verbal achievement accounted for by the mean score of the school, at grades 12 and 1 is as follows (Coleman, *et al.*, 1966, p. 296):

Group	Grade 12	Grade 1
Negro, South	22·54	23·21
Negro, North	10·92	10·63
White, South	10·11	18·64
White, North	7·84	11·07

Progressive Achievement Gap in a California School District

We searched for evidence of a PAG in our data in several ways, which can be only briefly summarized here. Separate analyses for each of the achievement tests did not reveal any striking differences in PAG, so the results can be combined without distortion of the essential results.

Mean Sigma Differences. The mean difference in sigma (standard deviation) units, based on the white group, by which Negro and Mexican-American pupils fall below the white group at each grade from 1 to 8 is shown in Table 3. The first three columns show the sample sizes on which the sigma differences are based. The sigma differences (i.e., σ below white mean) for Negroes and Mexican-Americans shown in columns 4 and 5 is the average of all the Stanford Achievement Tests given in each grade. Note that there is a reliable and systematic increase in the sigma difference from grade 1 to grade 3, for both Negro and Mexican groups, after which there is no further systematic change in achievement gap. The mean gap over all grades is ·66σ for the Negroes and ·55σ for the Mexicans. By comparison, look at columns 6 and 7, which show the mean sigma differences for those non-verbal ability tests in our battery which do not depend in any way upon reading skill and the content of which is not taught in school; this is the average sigma difference for the Lorge-Thorndike Non-verbal IQ, Figure Copying, and Raven's Progressive Matrices. We see that the sigma differences show a slight upward trend

Table 3: Number of white sigma units by which minority group means fall below the white mean

	Sample Size (N)			Stanford Achievement Tests		Non-verbal Intelligence		Home Index (SES)		Adjusted Achievement Means	
Grade	*White*	*Negro*	*Mexican*	*Negro*	*Mexican*	*Negro*	*Mexican*	*Negro*	*Mexican*	*Negro*	*Mexican*
1	285	218	258	·25	·34	1·07	·58	—	—	—·09	·15
2	229	162	250	·57	·37	1·09	·70	—	—	·15	·06
3	281	207	241	·83	·68	0·98	·53	·58	1·13	·11	·05
4	237	189	239	·69	·59	0·95	·48	·58	1·18	·17	·15
5	242	198	211	·75	·54	1·05	·62	·70	1·18	·21	·10
6	219	169	218	·84	·69	1·23	·67	·47	1·36	·09	·02
7	388	262	305	·71	·57	1·13	·72	·71	1·36	·07	·08
8	356	289	303	·64	·62	1·18	·79	·77	1·34	·06	·08
Mean				·66	·55	1·08	·63	·60	1·26	·10	·09

from the lower to the higher grades. Furthermore, the sigma differences are very significantly larger for the non-verbal intelligence tests than for the scholastic achievement tests in the case of Negroes (1·08σ for non-verbal intelligence *versus* 0·66 for achievement). The Mexicans show only a slight difference between their sigma decrement in non-verbal ability and in scholastic achievement (0·63 *versus* 0·55). If we can regard these non-verbal tests as indices of extra-scholastic learning ability, it appears then that these Negro children do relatively better in scholastic learning as measured by the Stanford Achievement Tests than in the extra-scholastic learning assessed by the non-verbal battery. In this sense, the Negro pupils, as compared with the Mexican pupils, are 'over-achievers', although the Negroes' absolute level of scholastic performance is 0·11σ below the Mexicans'. For the Negro group especially, the school can be regarded as an equalizing influence: Negro pupils are closer to white pupils in scholastic achievement than in non-scholastic non-verbal abilities. The mean Negro-white scholastic achievement difference is only 61 per cent as great as the non-verbal IQ difference. This finding is exactly the opposite of popular belief. The white *versus* Mexican achievement difference is 87 per cent as great as the non-verbal IQ difference.

Is there any systematic grade trend in our indices of socio-economic status and home environment? Columns 8 and 9 show the sigma differences below the white group on the composite score of Gough's Home Index, which assesses parental educational and occupational level, physical amenities, cultural advantages, and community involvement. (The Home Index was not used below grade 3.) There is a slight, but not highly regular, upward trend in these sigma differences for both Negro and Mexican groups, as if the students in the higher grades come from somewhat poorer backgrounds. Despite this, the sigmas for scholastic achievement (unlike the non-verbal ability tests) do not show any systematic increase from grade 3 to 8. Note also that on the Home Index the Mexicans, on the average, are further below the Negroes than the Negroes are below the whites. Moreover, the percentage of the Mexican children whose parents speak only English at home is 19·7 per cent as compared with 96·5 per cent for whites and 98·2 per cent for Negroes. In 14·2 per cent of the Mexican homes Spanish or other foreign language is spoken exclusively, as compared with 1·1 per cent for whites and 0·5 per cent for Negroes.

Covariance Adjustments of Achievement Scores. The next step of our analysis consists of obtaining covariance adjusted means on all the achievement tests, using all the ability tests,[1] along with sex and age in months, as the covariance controls. What this procedure shows, in effect, is the mean score on the achievement tests ('output') that would be obtained by the three ethnic groups if

[1]Lorge-Thorndike Verbal and Non-verbal IQ, Figure Copying, Raven's Matrices, Making X's, Listen Attention, and three memory tests.

they were equated on the ability tests ('input'). Although it is beyond the scope of this paper to explain in mathematical detail just how this kind of covariance adjustment is accomplished, a few words of explanation are in order to remove any mystery that may seem to exist for those who have not studied or used this statistical technique. A simplified illustration will give the reader some notion of what is involved.

The simplest possible illustration consists of two groups, say, Negro and white, who are given two tests, say, an IQ test and an achievement test. What we wish to find out is: what would be the mean achievement scores of the Negro and white groups if they were equated on IQ? What we must determine, in statistical terminology, is the 'covariance adjusted mean' achievement for each group. It is defined mathematically as

$$\dot{Y} = \bar{Y}_G - b(\bar{X}_G - \bar{\bar{X}}..)$$

In terms of our example,

$\dot{Y}_N$ = adjusted mean achievement score of Negro group

$\bar{Y}_N$ = raw mean achievement score of Negro group

$\bar{X}_N$ = mean IQ of Negro group

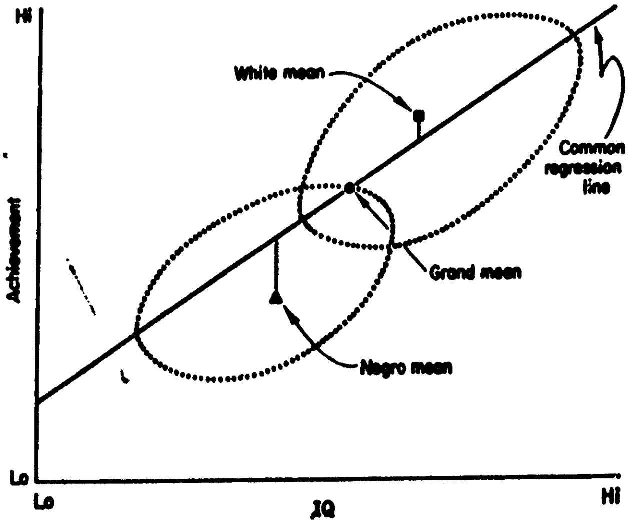
$\bar{\bar{X}}..$ = mean IQ of Negro and white groups combined, i.e., total mean IQ

b = the regression coefficient of Y on X, i.e., of achievement on IQ for both groups combined. The regression coefficient is the slope of the regression line. It is $r_{xy}\frac{\sigma_y}{\sigma_x}$, where, r_{xy} is the correlation between the two variables, X and Y (or IQ and achievement) and σ_x and σ_y are the standard deviations of these variables.

The situation can be pictured as follows.

For the sake of graphic clarity, this is a greatly exaggerated picture. The so-called regression line is the one straight line about which the squared deviations of all scores are a minimum. Thus, every individual score plays a part in determining the position and slope of the regression line. It is the one best-fitting line to the data of *all* the subjects in both groups. Although the mean raw achievement scores differ markedly for Negroes and whites in this illustration, we see that each group falls only slightly off the common regression line; in this example, the white mean is above the line and the Negro mean is below. The *adjusted* means for the two groups consist of the grand

Figure 2: Simplified correlation scatter diagram illustrating the regression of achievement on IQ and the covariance adjustment of hypothetical white and Negro achievement means



mean plus (or minus) the deviation of the particular group's mean from the regression line. If the means of both groups fall exactly on the common regression line, the adjusted means will be exactly the same and are equal to the grand mean. If there is zero correlation between the input (IQ) and output (achievement) variables, then the regression line will be perfectly horizontal and parallel to the base line, and the adjusted means will consequently be exactly the same as the raw (or unadjusted) means. In the above example, the white adjusted mean would be slightly higher than the Negro adjusted mean, because the white mean is above the regression line and the Negro below. The regression line can be thought of as predicting the most probable achievement score for any given IQ. If the correlation between IQ and achievement were perfect, one could predict achievement from IQ exactly, and vice versa.

The situation is essentially the same for adjusting the means of three or more groups, and one can easily picture another group placed in the above illustration. It is much more difficult to picture the situation when more than two variables are involved. In this illustration, we have one output variable (achievement) and only one input variable (IQ). It is possible to have two or three or more input variables. If there are two, then the situation would have to be pictured in three dimensions. The common regression line would no longer be a line on a two-dimensional surface but would become a plane in a three-dimensional cube, and we would be adjusting our means in terms of their deviations from the surface of this two-dimensional plane. If we go to three input variables the situation can no longer be pictured, since we would have to deal with a 'hyper plane' in four-dimensional space. Four input variables require a five-dimensional space, and so on. Although the problem can no longer be pictured graphically beyond two input variables, it can be solved mathematically for any number of input variables (although the point of diminishing returns is rapidly reached). For the sample sizes and the number of input variables used in the present study, the mathematical computations would be virtually impossible without the aid of a high speed computer.

Columns 10 and 11 of Table 3 show the sigma difference by which the Negro and Mexican covariance adjusted mean falls below that of the white group. These differences are quite small for both Negroes and Mexicans (averaging 0·10 and 0·09, respectively), and they show no systematic trend with grade level. In other words, when the minority groups are statistically equated with the majority (white) group on the ability test variables, their achievement, on the average, is less than 0·1 sigma below that of the white group. On an IQ scale that would be equivalent to 1·5 points, a very small difference indeed. The adjusted decrement is statistically significant, however, which raises the question of why it should differ significantly from zero at all. The reason could be actual differences between minority and majority schools in the effectiveness of instruction, or incomplete measurement of all the input variables relevant to scholastic learning, or some lack of what is called homogeneity of regression for the three ethnic groups, which works against the covariance adjustment. We know the latter factor is involved to some extent, and some combination of all of them are most likely involved. But taken all together, the fact that the majority-minority difference in mean adjusted achievement scores is still less than $0{\cdot}1\sigma$ means the direct contribution of the schools to the difference must be even smaller than this, if existent at all. Surely it is of practically negligible magnitude.

When the personality variables (the Junior Eysenck Personality Inventory) and the four scales of the Home Index are also included with the ability variables in obtaining covariance adjusted means, the ethnic differences in scholastic achievement are wiped out almost entirely. Two-thirds of the majority-minority differences (for various achievement sub-tests at various grades) are not significant at the five per cent level and are less than $0{\cdot}1\sigma$. The adjusted mean differences *between* ethnic groups are smaller than the grade-to-grade sigma differences *within* ethnic groups. From this analysis, then, the school's contribution to ethnic achievement differences must be regarded as nil. If the input variables themselves are strongly influenced by the school to the disadvantage of the minority children, we should expect to find a greater sigma difference for non-verbal IQ at grade 8 (age 13) than at

kindergarten. In the present study Negroes are $1{\cdot}11\sigma$ below whites in non-verbal IQ in kindergarten as compared with $1{\cdot}17\sigma$ in grades 7 and 8 —a trivial difference. Mexican children are $0{\cdot}98\sigma$ below whites in non-verbal IQ at kindergarten and $\cdot 88\sigma$ below at grades 7 and 8. Thus the minority children begin school at least as far below the majority children in non-verbal ability as they are by grades 7 and 8. The schools have not depressed the ability level of minority children relative to the majority, but neither have they done anything to raise it. Differences in verbal IQ are slightly more likely to reflect the effects of schooling, and we note that in grades 7 and 8 Negroes are $1{\cdot}00\sigma$ below the white mean and Mexicans are $0{\cdot}90\sigma$ below.

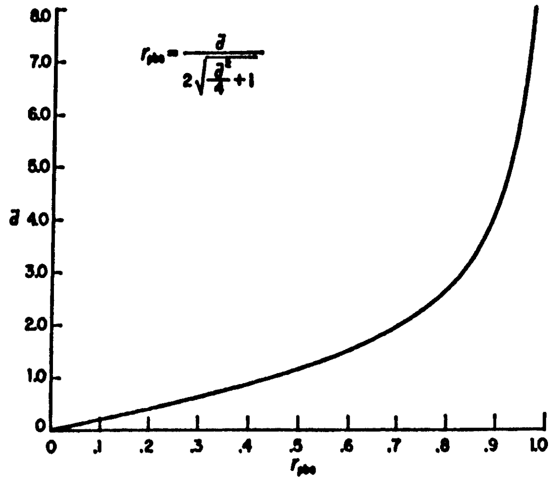
Paired Ethnic Group Differences. The maximum discrimination that we can make between the three ethnic groups in terms of all of our 'input' variables (ability tests, personality inventories, and socio-economic indexes) is achieved by means of the multiple point-biserial correlation coefficient. The product-moment correlation obtained between a continuous variable (e.g., IQ) and a quantized (dichotomous) variable (e.g., male *versus* female, where male = 1 and female = 0) is called a point-biserial correlation (r_{pbi}). Mathematically it is defined as:

$$r_{pbi} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_t}\sqrt{pq}$$

where $\bar{X}_2$ and $\bar{X}_1$ = means of groups 1 and 2
σ_t = standard deviation of total (i.e., groups 1 and 2 combined)
p and q = proportions of total sample in groups 1 and 2, respectively ($p + q = 1{\cdot}00$)

It is also possible to compute r_{pbi} in the same manner that one computes the Pearson product-moment correlation between any two continuous variables, except that the dichotomous variable is quantized by assigning 0 and 1 to its two categories. It is also possible to obtain a multiple point-biserial correlation, which gives the maximum possible correlation between the quantized variable and the best weighted combination of a number of 'predictor' variables. The multiple correlation thus represents the maximum degree of discrimination that can be achieved between the two categories of the quantized variable by means of the particular set of predictor variables.

Figure 3: The relationship between the point-biserial correlation (r_{pbi}) and the mean difference ($\bar{d}$) between groups in sigma units on the continuous variable, assuming equal sigmas and equal Ns in the two groups



Since the multiple correlation capitalizes upon sampling error (chance deviations from population values) to achieve the maximum value of the correlation, it is spuriously inflated by a degree that is inversely proportional to the sample size and the number of variables correlated. For this reason, the obtained multiple correlation should be 'shrunk' down to its estimated population value (i.e., its value if there were no sampling error). The method for doing this is given in most statistics textbooks (e.g., Guilford, 1956, pp. 398-399). All the multiple correlations reported here have thus been 'shrunk' and therefore represent a conservative estimate of the amount of discrimination achieved between the ethnic groups by our battery of 'input' tests.

When the sizes of the samples entering into the quantized variable are large and nearly equal, and when they have nearly equal standard deviations on the predictor variables, it is possible roughly to 'translate' the point-biserial correlation into a linear mean distance in constant sigma units between the two categories of the quantized variable. Figure 3 shows the function relating the point-biserial correlation to the mean sigma difference ($\bar{d}$) between groups. The r_{pbs} can attain a value of 1·00 only if the variance within each group diminishes to zero.

Table 4 gives the multiple point-biserial correlations between each ethnic dichotomy and all the 'input' variables—first just the ability tests and second the ability tests plus the personality inventory and socio-economic index. Note that the three groups are almost equally discriminable from one another in terms of the multiple correlation, especially after the personality and social background variables are added to the predictors. This is interesting, because it means that the two minority groups, though both are regarded as educationally and socio-economically disadvantaged, actually differ from one another on this composite of all input variables almost as much as each one differs from the majority group. The Negro and Mexican groups each differ from the majority group in a somewhat different way in terms of total pattern of scores, and they differ from one another almost as much. A factor analysis, shown in the next section, helps to reveal the ways in which the three groups differ from one another.

The last three columns in Table 4 show the correlation between each ethnic dichotomy and the Stanford Achievement Tests, with all the 'input' variables partialled out, i.e., statistically held constant. These correlations represent the average contribution made to the ethnic discrimination by the Stanford Achievement Tests regarded independently of the 'input' variables. It can be seen that these correlations are very small indeed. For the sample sizes used here,

Table 4: Point-biserial multiple correlations for 'input' variables and partial correlation for 'output' with 'input' held constant

	'INPUT' *All Ability Tests*			'INPUT' *Ability + Personality + Home Index*			'OUTPUT' *Stanford Achievement Minus All 'Input' Variables*		
Grade	W-N	W-M	M-N	W-N	W-M	M-N	W-N‡	W-M	M-N
1	·49	·28	·29	—	—	—	—	—	—
2	·54	·47	·37	—	—	—	—	—	—
3	·54	·45	·35	·62	·59	·46	·06	·02	·07
4	·48	·38	·41	·55	·60	·55	·15	·07	·09
5	·47	·38	·27	·60	·59	·36	·13	·05	·11
6	·53	·47	·42	·69	·67	·59	·14	·11	·04
7	·52	·42	·26	·68	·70	·45	·09	−·04	·11
8	·57	·42	·43	·65	·66	·46	·06	−·02	−·07
Mean	·52	·41	·36	·63	·64	·48	·11	·05	·07

†Partial correlations of less than 0·10 are not significant at the five per cent level.
‡The quantised ethnic groups are White = 3, Mexican = 2, Negro = 1, so that for W-N and W-M positive correlations indicate higher achievement scores for the white group, and a positive correlation for M-N indicates higher scores for the Mexican group.

correlations of less than 0·10 can be regarded as statistically non-significant at the five per cent level. The proportion of the total variance between the ethnic groups that is accounted for by the achievement tests is represented by the square of the correlation coefficient. Applied to the partial correlations for the Achievement Tests in Table 4, this shows how trifling are the ethnic group achievement differences after the ethnic group differences on the input variables have been controlled.

Factor Analysis of All Variables. A factor analysis (varimax rotation of the principal components having Eigenvalues greater than 1) was carried out at each grade level on all test variables obtained at that grade level plus three others: sex, chronological age in months, and welfare status of the parent (whether receiving welfare aid to dependent children). The latter variable was added to supplement the indices of socio-economic status (the four scales of Gough's Home Index). Since grades 4, 5 and 6 had all the measures (27 variables) and the same tests were used at each of these grades, they are the most suitable part of our total sample for factor analytic comparisons. The results are essentially the same at all grade levels, although because the personality inventory and the Home Index were not used in the primary grades, and the Figure Copying Test was not used beyond grade 6, not all of the factors that emerged at grades 4, 5 and 6 come out at one or another of the other grades. Moreover, because of the large number of variables entering into the analysis at grades 4-6, more small factors come out which, in a sense, 'purify' the main factors by partialling out other irrelevant and minor sources of variance.

Factor analyses were performed first on the three ethnic groups separately to determine if essentially the same varimax factors emerged in each group. They did. All three groups yield the same factors, with only small differences in the loadings of various tests. This finding justifies combining all three groups for an overall factor analysis of the total student sample at each grade level. This was done. Eight factors with Eigen-

Table 5: Loadings of variables on first principal component for grades 4 to 8 (decimals omitted)

Variable	Grade 4	5	6	7	8
1. Sex (M — 0, F — 1)	14	14	03	08	12
2. Extraversion	25	28	46	33	24
3. Neuroticism	00	—06	—21	—12	01
4. Lie Scale	—17	—11	—19	—27	—39
5. Home Index—1	31	45	41	49	48
6. Home Index—2	29	30	34	41	45
7. Home Index—3	36	41	27	50	44
8. Home Index—4	29	43	28	47	40
9. Aid to Dependent Children	—21	—43	—32	—31	—26
10. Age in Months	—05	—09	—04	—04	—12
11. Lorge-Thorndike Verbal IQ	85	88	85	88	87
12. Lorge-Thorndike Non-verbal IQ	73	75	76	79	83
13. Raven's Progressive Matrices	54	55	54	54	63
14. Figure Copying	45	51	57	—	—
15. Listening-Attention	11	19	21	06	12
16. Memory—Immediate	45	40	36	27	32
17. Memory—Repeat	44	33	24	25	27
18. Memory—Delayed	43	41	41	25	27
19. Making X's 1st Try	14	02	31	53	10
20. Making X's 2nd Try	19	14	29	48	19
21. SAT: Word Meaning	83	81	81	—	—
22. SAT: Paragraph Meaning	80	79	89	86	83
23. SAT: Spelling	75	76	78	73	73
24. SAT: Language	83	84	87	78	75
25. SAT: Arithmetic Computation	57	45	63	73	73
26. SAT: Arithmetic Concepts	72	62	80	76	83
27. SAT: Arithmetic Applications	77	71	82	72	71
Per Cent of Variance	22	26	29	28	27

values greater than 1 emerged at grades 4, 5 and 6, accounting respectively for 67 per cent, 66 per cent and 70 per cent of the total variance.

The first principal component can be regarded as the general or *g* factor for this set of 27 variables. Table 5 shows the loadings of each of the 27 (or 25 in grades 7 and 8) variables on the first principal component in grades 4 to 6. The first principal component is the single most general factor accounting for more of the variance than any other factor. It is most heavily loaded in the Stanford Achievement Tests and Verbal IQ. Inspection of the loadings of the other variables gives an indication of their correlation with this most general achievement factor.

The eight principal components were rotated to approximate simple structure by the varimax criterion. In grades 4, 5 and 6 four substantial and clear-cut factors emerged. The remaining factors serve mainly to pull out irrelevant variance from the main factors. The four main factors that emerge are:

Factor I. Scholastic Achievement and Verbal Intelligence

Variables	Factor Loading *Gr. 4*	*Gr. 5*	*Gr. 6*
Lorge-Thorndike Verbal IQ	·75	·75	·85
Word Meaning	·83	·69	·82
Paragraph Meaning	·83	·77	·89
Spelling	·82	·77	·81
Language	·82	·79	·86
Arithmetic Computation	·64	·58	·65
Arithmetic Concepts	·73	·69	·83
Arithmetic Applications	·77	·71	·85

Factor II. Non-verbal Intelligence

Variables	Factor Loading *Gr. 4*	*Gr. 5*	*Gr. 6*
Lorge-Thorndike Non-verbal IQ	·61	·57	·32
Raven's Progressive Matrices	·75	·75	·55
Figure Copying	·69	·68	·41

Factor III. Rote Memory Ability

Variables	Factor Loading *Gr. 4*	*Gr. 5*	*Gr. 6*
Memory Span—Immediate Recall	·85	·81	·77
Memory Span—Repeated Series	·85	·81	·86
Memory Span—Delayed Recall	·83	·79	·74

Factor IV. Socio-economic Status

Variables *Home Index:*	Factor Loading *Gr. 4*	*Gr. 5*	*Gr. 6*
1. Parental Education and Occupation	·75	·74	·77
2. Physical Amenities	·69	·77	·72
3. Community Participation	·66	·76	·75
4. Cultural Advantages	·66	·59	·66
Receives Welfare Aid to Dependent Children	—·40	—·34	—·46

The remaining four minor factors are: (1) Speed, motivation, persistence as defined principally by the Making X's Test, (2) Neuroticism, (3) Extraversion, (4) Age in months. These variables, having their largest loadings on separate factors, are in effect partialled out of the major factors. The four major factors listed above are orthogonal, i.e., uncorrelated with one another, and each one is thus viewed as a 'pure' measure of the particular factor in the sense that the effects of all the other factors are held constant.

Table 6: Mean varimax factor scores for three ethnic groups in grades 4, 5 and 6

			Mean Factor Scores							
			I Verbal IQ and Achievement		II Non-verbal IQ		III Memory		IV Socio-Economic Status	
Grade	*Group*	*N*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
4	White	113	55·2	10·7	51·6	8·1	51·6	9·4	53·8	10·3
	Negro	129	47·1	6·5	44·6	8·9	51·0	11·2	51·7	7·9
	Mexican	145	49·5	8·3	51·0	9·3	48·1	7·7	43·6	7·8
5	White	144	54·7	8·7	52·3	8·2	50·4	9·1	54·1	9·2
	Negro	132	45·5	8·4	47·0	11·1	51·1	9·9	49·7	9·5
	Mexican	135	49·6	8·5	50·1	8·5	48·2	9·5	44·6	8·1
6	White	131	55·0	8·8	50·9	7·2	50·7	8·8	53·8	9·4
	Negro	124	47·1	8·3	44·1	10·5	50·5	9·9	51·5	8·0
	Mexican	126	49·1	9·3	51·0	8·7	48·0	10·2	42·5	7·5

Ethnic Group Comparisons of Factor Scores. The final step was to obtain factor scores for every student on each of these four main factors. For the total sample, within each grade, these factor scores are represented on a T-score scale, i.e., they have an overall mean of 50 and a standard deviation of 10. Table 6 shows the mean and standard deviation of the factor scores for each of the ethnic groups.

Note that the ethnic group differences in Factor I do not show any systematic increase from grade 4 to 6, thus lending no support to the existence of a cumulative deficit in the minority groups. Analysis of variance was performed on the factor scores and Scheffé's method of contrasts was used for testing the statistical significance of the differences between the means of the various ethnic groups at each grade level. The results of these significance tests are shown in Table 7. We see that in Factor I (Verbal IQ and Scholastic Achievement) both minority groups are significantly below the majority group, and Negroes are significantly below the Mexican group except in grade 6, where the difference is in the same direction but falls short of significance.

On Factor II (Non-verbal Intelligence) Negroes fall significantly below whites and Mexicans at all grades, and the differences between Mexicans and whites are non-significant at all grades. It should be remembered that this non-verbal intelligence factor represents that part of the variance in the non-verbal tests which is not common to the verbal IQ and achievement tests or to the memory tests. The Mexican-white difference is significant on that part of the ability tests variance which has most in common with scholastic achievement and is represented in Factor I.

Factor III (Rote Memory) shows no significant differences between the Negro and white groups; the Mexican group is significantly below the white at grade 4 and below the Negro at grades 4 and 5. This finding is consistent with the findings of other studies that mean differences between groups of lower and middle socio-economic status are smallest on tests of short-term memory and rote learning (Jensen, 1968b).

Factor IV (socio-economic status) shows relatively small differences between the Negro and white groups, while the Mexican group is significantly below the other two. Again, it should be realized that we are dealing here with 'pure' factor scores which are independent of all the other variables. Thus Factor IV shows us the relative standing of the three ethnic groups in socio-economic status when all the other variables are held constant. What these results indicate is that Negro and white children statistically equated for intelligence, achievement, and memory ability differ very little in socio-economic status as measured by our indices, but that Mexican children, when equated on all other variables with white children or with Negro children, show a comparatively much poorer

Table 7: The significance of ethnic group differences in mean factor scores, by Scheffé's method of contrasts

		Factors			
Contrasts (Means)	Grade	I Verbal IQ and Achievement	II Non-verbal Intelligence	III Memory	IV Socio-economic Status
Negro-White	4	—**	—**	—NS	—NS
	5	—**	—**	+NS	—**
	6	—**	—**	—NS	—NS
Mexican-White	4	—**	—NS	—*	—**
	5	—**	—NS	—NS	—**
	6	—**	+NS	—NS	—**
Mexican-Negro	4	+*	+**	—*	—**
	5	+**	+*	—*	—**
	6	+NS	+**	—NS	—**

*p <0·05
**p <0·01

NS — Not Significant

background than either the white or Negro groups. On the present measures, at least, the Mexicans must be regarded as much more environmentally disadvantaged than the Negroes, and this takes no account of the Mexican's bilingual problem. In view of this it is quite interesting that Mexican pupils on the average significantly exceed the Negro pupils in both verbal and non-verbal intelligence measures and in scholastic achievement.

Equality of educational opportunity: uniformity or diversity of instruction?

The results of our analysis thus far fail to support the hypothesis that the schools have discriminated unfavourably against minority pupils. When minority pupils are statistically equated with majority children for background and ability factors over which the schools have little or no control, the minority children perform scholastically about as well as the majority children. The notion that poor scholastic achievement is partly a result of the pupil's ethnic minority status *per se*, implying discriminatory schooling, is thus thoroughly falsified by the present study. This does not imply that the same results would be obtained in every other school system in the country. Where true educational inequalities between majority and minority pupils exist, we should expect the present type of analyses to reveal these inequalities, and it would be surprising if they were not found in some school systems which provide markedly inferior educational facilities for minority pupils. It should be noted, on the other hand, that the present study was conducted in a school district which had taken pains to equalize educational facilities in schools that serve predominantly majority or predominantly minority populations. The success of this equalization is evinced in the results of the present analyses.

But we can take a bold step further and ask: Is equalization of educational facilities enough? Is the real meaning of equality of educational opportunity simply uniformity of facilities and instructional programmes? Is it possible that true equality of opportunity could mean doing whatever is necessary to maximize the scholastic achievement of children, even if it might mean doing quite different things for different children in terms of their differing patterns of ability? Note that I did not say in terms of their ethnic or social class status, but in terms of their individual patterns of ability. The fact that different social classes and ethnic groups show different modal patterns of ability, of course, means that different proportions of various sub-populations will have different patterns of strength and weakness in various mental abilities. Is such a fact to be deplored and swept out of sight, or should it be examined with a view to utilizing the differences in the design of instructional programmes that might maximize each individual's benefits from schooling? A couple of years ago I wrote: 'If we fail to take account either of innate or acquired differences in abilities and traits, the ideal of equality of educational opportunity can too easily be interpreted so literally as to be actually harmful, just as it would be harmful for a physician to give all his patients the same medicine. One child's opportunity can be another's defeat' (Jensen, 1968a, p. 3). At that time I suggested that we look for differential ability patterns that might interact with different instructional methods in such a way as to maximize school learning for all individuals and at the same time minimize individual and group differences in scholastic achievement and any other benefits derived from schooling.

In our laboratory research we have discovered two broad classes of abilities which show marked differences in their relation to social class and race (Jensen, 1968b, 1969b, 1970a; Jensen and Rohwer, 1968, 1970). Briefly, what we have found is that children of low socio-economic status, especially minority children, with low measured IQs (60 to 80) are generally superior to middle-class children with similar IQs in tests of associative learning ability—free recall, serial rote learning, paired-associates learning, and digit span memory. This finding has been interpreted theoretically in terms of a hierarchical model of mental abilities, going from associative learning to conceptual thinking, in which the development of lower levels in the hierarchy is necessary but not sufficient for the development of higher levels. Our hypothesis states that the continuum of tests going from associative to conceptual is the phenotypic expression of two functionally dependent but genotypically inde-

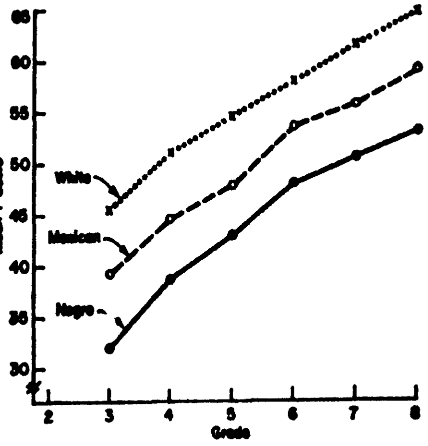
pendent types of mental processes, which we call Level I and Level II. Level I processes are perhaps best measured by tests such as digit span and serial rote learning; Level II processes are represented in tests such as the Progressive Matrices. Level I and Level II abilities are distributed differently in upper and lower social classes and in different ethnic groups. Level I is distributed fairly evenly in all sub-populations. Level II, however, is distributed about a higher mean in upper than in lower social classes. The majority of children now called 'culturally disadvantaged' show little or no deficiency in Level I ability but are about one standard deviation below the general population mean on tests of Level II ability. Children who are above average on Level I but below average on Level II ability usually appear to be bright and capable of normal learning and achievement in many life situations, although they have unusual difficulties in school work under the traditional methods of classroom instruction. Many of these children, who may be classed as retarded in school, suddenly become socially adequate persons when they leave the academic situation. But children who are below average on both Level I and Level II seem to be much more handicapped. Not only is their scholastic performance poor, but their social and vocational potential also seem to be much less than those of children with normal Level I functions. Yet both types of children look much alike in overall measures of IQ and scholastic achievement.

These findings are important because they help to localize the nature of the intellectual deficit of many children called 'culturally disadvantaged'. We must ask whether we can discover or invent instructional methods that engage Level I more fully and thereby provide a means of improving the educational attainments of many of the children now called 'culturally disadvantaged'. In our current instructional procedure are we utilizing so exclusively those mental abilities we identify as IQ (Level II) that children who are relatively low in IQ but have strength in other abilities are unduly disadvantaged in the traditional classroom? The whole complex process of classroom instruction as we know it has evolved in relation to a relatively small upper-class segment of Anglo-European stock. The modal pattern of development in learning abilities of this group has probably shaped to a considerable degree the particular educational procedures public education has long regarded as standard for everyone, regardless of differences in cultural background or inherited patterns of ability. But so far we have not successfully met the challenge presented by our ideal of a rewarding education for all segments of the population, with their diverse patterns of ability.

Looking, for example, at the factor scores shown in Table 6 we note that the minority groups are not significantly below the majority group on Factor III (Memory), which we would identify with Level I ability. Lest anyone try to argue that these 'pure' factor scores do not correspond to any 'impure' scores that could be obtained with actual tests, we can look at Figures 4 and 5, showing the grade-to-grade growth curves of a good Level II test (Raven's Progressive Matrices) and a good Level I test (a composite of the three digit memory tests).

The results of both tests have been put on the same scale of T scores, with an overall mean of 50 and a standard deviation of 10 (based on the standard deviation of raw scores in the white group at grade 5). The differences between the

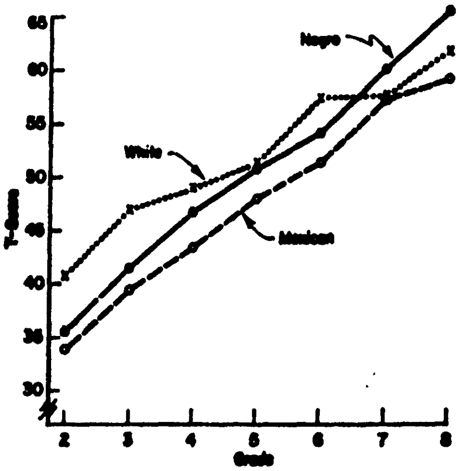
Figure 4: Mean T scores ($\bar{X}$ = 50, SD = 10) on Raven's Progressive Matrices in Grades 3 to 8



growth curves shown in Figures 4 and 5 are striking. The approximately one standard deviation difference between the Negro and white groups on the Level II test (Matrices) can be seen to have rather drastic implications in terms of grade level comparisons. By drawing a horizontal line from the Negro or Mexican mean at any grade to the point where it crosses the curve for the white group and dropping a perpendicular to the baseline, we can read off the grade equivalent of the minority group mean. The average Negro eighth-grader in this school system, for example, performs on the matrices at a level equivalent to white children at grade 4.5. Mexican children at grade 8 perform at grade 6.3. The grade 6 performance of Negroes and Mexicans is equivalent to the white's performance in grades 3.4 and 4.5, respectively.

On the other hand, note the small differences between the groups on the Level I test (Memory Span) in Figure 5. It is interesting to conjecture whether instruction in scholastic skills specifically aimed at Level I ability in children who are low in Level II would significantly reduce majority-minority differences in scholastic achievement. We do not know and can find out only through further research. If instruction is aimed only at Level II ability for all children, we should expect

Figure 5: Mean *T* scores ($\bar{X}$ = 50, *SD* = 10) on composite memory score in Grades 2 to 8



sizeable majority-minority differences in achievement. If instruction could somehow be aimed at Level I ability for all those children (regardless of ethnic identification) who are significantly stronger in Level I than in Level II, would their achievement be brought appreciably closer to that of the majority? Or is scholastic learning so intrinsically dependent on Level II ability that no form of instruction attempting to capitalize on Level I ability could possibly succeed beyond the most elementary aspects of any academic subject matter? Again, we do not know. But until these possibilities are explored, schools may be accused of cheating many children, especially large numbers of minority children, by providing uniform facilities but not sufficiently diversified instructional programmes to minimize differences in achievement and also maximize the overall level of achievement.

Some scholastic subjects would seem to lend themselves more to Level I processes and instructional methods than other subjects. For instance, the learning of spelling and arithmetic computation would seem to be less dependent upon Level II ability than, say, reading comprehension, arithmetic concepts or arithmetic applications. If this is true, we should expect majority-minority differences to be smaller on the Level I types of subject matter than on the Level II types. Let us make the relevant comparisons in the data of the present study. Table 8 shows these comparisons in sigma units. They bear out our hypothesis; the pupils of both minority groups fall below the majority mean about one-fifth of a sigma more on Level II-like scholastic achievement than on Level I-like subjects. Clearly, school subjects which by their nature

Table 8: Mean sigmas (based on white group) below white mean of Negro and Mexican pupils in grades 4-8 on level I-like and level II-like tests of scholastic achievement

Tests	Negro (N=1, 107)	Mexican (N=1, 276)
Level I-Like Tests:		
Spelling	·62	·32
Arithmetic Computation	·56	·36
Level II-Like Tests:		
Paragraph Meaning	·90	·75
Arithmetic Concepts	·71	·60
Arithmetic Applications	·72	·55

77-942 O - 72 - 18

seem to permit greater utilization of Level I ability show smaller majority-minority differences than those subjects which involve more Level II ability. This raises the interesting question whether all scholastic subjects can be taught in ways that *maximize* their dependence on Level I and *minimize* their dependence on Level II. If this can be done for children who are low in Level II ability—and we will never know without trying—it should reduce not only the scholastic achievement gap between majority and minority children but the achievement differences among all children of every group. If it succeeds, it would do so, not by pulling anyone down toward the common average, but by capitalizing on each child's particular strengths and minimizing the role of his particular weaknesses in learning any given kind of subject matter. This would seem to be an avenue worth exploring in our efforts to achieve not only equality of educational opportunity but greater equality of scholastic performance as well.

References

BOWLES, S. and LEVIN, H. M. (1968). 'The determinants of scholastic achievement—an appraisal of some recent evidence', *J. Hum. Resources*, Winter.

BURT, C. (1961). 'The gifted child', *Brit. J. Statist. Psychol.*, 14, 123-39.

COLEMAN, J. S. *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: US Office of Education.

GOUGH, H. G. (1949). 'A short social status inventory', *J. Educ. Psychol.*, 40, 52-6.

GUILFORD, J. P. (1956). *Fundamental Statistics in Psychology and Education* (3rd ed.). New York: McGraw-Hill.

ILG, F. L. and AMES, L. B. (1964). *School Readiness: Behavior Tests Used at the Gesell Institute*. New York: Harper & Row.

JENSEN, A. R. (1967). 'Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins', *Proceedings of the National Academy of Sciences*, 58, 149-57.

JENSEN, A. R. (1968a). 'Social class, race, and genetics: implications for education', *Amer. Educ. Res. J.*, 5, 1-42, (a).

JENSEN, A. R. (1968b). 'Patterns of mental ability and socio-economic status', *Proceedings of the National Academy of Sciences*, 60, 1330-7, (b).

JENSEN, A. R. (1969a). 'How much can we boost IQ and scholastic achievement?', *Harv. Educ. Rev.*, 39, 1-123, (a).

JENSEN, A. R. (1969b). 'Intelligence, learning ability, and socio-economic status', *J. Special Educ.*, 3, 23-35, (b).

JENSEN, A. R. (1970a). 'A theory of primary and secondary mental retardation'. In: ELLIS, N. R. ed. *International Review of Research in Mental Retardation*, Vol. IV. New York: Academic Press.

JENSEN, (1970b). A. R. 'Hierarchial theories of mental ability'. In: DOCKRELL, B. ed. *On Intelligence*. London: Methuen.

JENSEN, A. R. and ROHWER, W. D., jr. (1970a). 'Mental retardation, mental age, and learning rate', *J. Educ. Psychol.*, 59, 402-403.

JENSEN, A. R. and ROHWER, W. D., jr. (1970b). *An Experimental Analysis of Learning Abilities in Culturally Disadvantaged Children*. Final Report, Office of Economic Opportunity. Contract No. OEO 2404.

THORNDIKE, R. L. (1951). 'Community variables as predictors of intelligence and academic achievement', *J. Educ. Psychol.*, 42, 321-38.

WECHSLER, D. (1958). *The Measurement and Appraisal of Adult Intelligence*. (4th ed.). Baltimore: Williams & Wilkins.

WILSON, A. B. (1967). 'Educational consequences of segregation in a California community'. *Racial isolation in the public schools*, Appendices, Vol. II, p. 185. US Commission on Civil Rights, Washington, DC.

WISEMAN, S. (1964). *Education and Environment*. Manchester: Manchester University Press.

(*Manuscript received 14th January 1971*)

ENVIRONMENTALIST RATIONALIZATION VERSUS ENVIRONMENTAL RESEARCH

By Arthur R. Jensen

In the discussion that follows, it is necessary to distinguish "environmentalism" from research on the environment. Environmentalism is the scientifically anomalous attitude that ignores, shuns, or denigrates any hypothesis of genetic causation in specific classes of human individual or group differences. Environmentalists differ among themselves in the kinds of differences from which they dogmatically exclude the possibility of genetic influences. Thus we see environmentalists who accept the findings on the heritability of individual differences in intelligence but who vehemently argue against the suggestion that genetic factors may be involved in any subpopulation differences, social-class or racial. Still others acknowledge the evidence on genetic intelligence differences among social classes within racial groups, but categorically reject without evidence the hypothesis that specific racial groups differ genetically in mental abilities. Some will admit genetic explanations, or at least grant their plausibility, regarding racial differences in physical and sensory capacities, while not

allowing the possibility of genetic differences in more complex mental capabilities. The idea that certain small and isolated racial groups, such as the Australian Bushmen, might differ genetically from major racial groups in mental capacities is viewed only with a mild skepticism by some environmentalists who vociferously denounce anyone who questions wholly environmental theories of intelligence differences between major racial groups.

The aim of the environmentalist, as a matter of principle, is to "explain" a given human difference as due wholly to environmental causes -- no matter what. As already noted, environmentalists often differ in the paritcular kinds of traits and groups to which they extend their emotional need for a wholly environmental rationalization of human differences. This tendency results in the uncritical acceptance of almost any environmental factor that anyone suggests as an explanation, regardless of its often purely *ad hoc* status, its inconsistency with other data, and often the failure even to show any *correlation* (much less causation!) between the suggested environmental causes and the behavioral traits in question. Since an environmental explanation is decreed as necessary and sufficient, almost any environmental factor will do, without the need to demonstrate its causal connection, or even correlation, with intelligence or scholastic achievement. Some environmental factors are formulated clearly enough to be put to the test of evidence; as each of the hypothesized factors is rejected on the basis of evidence, other increasingly subtle environmental deficits are postulated to explain the differences. Baratz and Baratz (1970, p. 35) have noted this tendency in various attempts to account for the failure of intervention programs such as Head Start to appreciably raise IQ and scholastic performance: "Postulation of one deficit which is unsuccessfully dealt with by intervention programs then leads to the

discovery of more basic and fundamental deficits. Remediation or enrichment gradually broadens its scope of concern from the fostering of language competence to a broad-based restructuring of the entire cultural system. The end result of this line or argument occurs when investigators such as Deutsch and Deutsch (1968) postulate that 'some environments are better than others.'"

Inconsistencies abound in environmentalist arguments. Unrelated children adopted at birth and reared together are much less alike in IQ' than true siblings, it is said, because of subtle factors within the family environment which makes them dissimilar in intelligence. In the next breath it is argued that identical twins reared apart in different families are highly similar in IQ because of subtle influences common to both families (though they may be at opposite ends of the SES spectrum and have no knowledge of one another) which make for a high correlation between the twins' IQs. How often do we see environmentalists propose any experiment, statistical study, type of evidence, or any combination thereof, which could cause them to question the null hypothesis regarding genetic differences, or even to reject a particular enviornmental factor which has been postulated as a cause of IQ differences? Various environmental factors are constantly repeated in the environmentalist literature as a cause of IQ differences, even when studies specifically designed to test these hypotheses have yielded nothing but negative results. There seems to be no way for the environmentalist to give up any hypothesized environmental factor; regardless of the outcomes of empirical tests, each newly hypothesized factor is added to the growing list of purported environmental causes of IQ differences.

The principal environmentalist fallacy consists of looking for any environmental differences that exist between two subpopulation groups which differ in mean IQ and merely <u>assuming</u> that the environmental differences

are the cause of the IQ difference; usually it is not even regarded as necessary to demonstrate that a non-zero correlation exists between the hypothesized environmental factor and IQ. At least three critical questions need to be answered about every hypothesized environmental factor before one can even begin to consider whether it is a causal factor: (1) does it correlate with the trait in question within the two groups being compared? (2) how much do the groups differ on the environmental factors? (3) does the factor make any significant contribution to within groups or between groups variance in the trait independently of other hypothesized factors? As pointed out in a previous section, one cannot properly assess the importance of a large number of intercorrelated environmental factors from the single (zero order) correlations of each one with IQ. It is each variable's independent contribution to the multiple correlation that counts. When major environmental factors fail to account for IQ differences sufficiently to sustain the environmental hypothesis, other subtler environmental factors are then postulated, and they may be tested in a new study and be found to show some correlation with the IQ difference. But rarely are they combined with the variables of the first study to see if they in fact add any significant increment to the coefficient of multiple determination ($\underline{R}^2$). Even if they do, the direction of the causality often remains an open question which can be answered only by other evidence than correlational data. The pattern of multiple correlations at best narrows the range of possibilities in seeking the most probably fruitful environmental variables for experimental manipulation.

It is this approach which distinguishes research on environmental factors, which is a legitimate scientific enterprise, from environmentalism, which is a religious dogma. Behavior geneticists recognize the influence of non-genetic factors in all forms of behavior and in all individual and

group differences, and they are interested in understanding these non-genetic factors precisely and in learning what proportions of the variance they contribute, singly and in combination, and how much is due to additive, interactive, and covariance effects in the population. Developmental behavior genetics seeks to understand how the individual phenotype develops through the genotype's interaction with and utilization of the environment. Variance in genotypes for any trait, within or between groups, is not ruled out or restricted on any *a priori* basis. Environmentalists, on the other hand, simply decree the null hypotheses *a priori* with respect to certain classes of genetic variance, and in order to fill the void must posit a number of environmental influences or measurement biases which often are accepted merely on the grounds of plausibility. It is possible, however, to bring research evidence to bear on many specific environmental factors and test biases which environmentalists assume are the main causes of subpopulation differences in intelligence and related performance. Under such examination of the relevant evidence, some of the main pillars of the environmentalist argument regarding Negro-white intelligence differences simply collapse, still others are seen to be resting on extremely flimsy foundations in fact, while the remainder are so vaguely formulated as to be insusceptible to empirical proof or disproof.

Equating for Socioeconomic Variables

In comparative studies of the mental abilities of racial groups, environmentalists are most insistent that the racial samples being compared on intelligence be matched, or otherwise equated, on indices of socioeconomic status (SES), which usually includes father's occupation, education of parents, income, quality of housing, and place of residence. When groups are thus "equated" and a substantial mean IQ difference still shows up, it is

claimed that not enough environmental factors were controlled. As one sociologist put it: "...the kinds of socioeconomic measures that have been used so far in attempting to control on environmental effects appear to omit a wealth of cultural and psychological factors." This is a testable hypothesis; it should be determined how much the cultural and psychological factors (assuming they can be specified and measured) add to the multiple $\underline{R}^2$ with IQ over and above the $\underline{R}^2$ yielded by good indices of SES.

But the whole notion of equating for SES, in the first place, involves what has been called the "sociologist's fallacy." This fallacy is seen in full bloom in one sociologist's crticism of studies of Negro-white IQ differences which equated the groups for SES or other environmental factors: "Actually in most of the studies he {Jencen, 1969a} reports on, the most important environmental variable, the IQ of the parent, has not been equated at all" (Stinchcombe, 1969, p. 516). Aside from the strictly environmental effect of parental IQ,[69*] it is obvious that, since IQ variance contains a large genetic component, equating groups for parental IQ means equating them for genetic factors more than for environmental factors. The same is true, though to a lesser degree, when we equate for SES. When typical Negro children are equated with white children on some index of SES, one is comparing a majority of the Negro population with some lower fraction (between one-fifth and one-sixth) of the white population.[70] The white comparison group, therefore, is not genetically representative of the entire white population but is genotypically (as well as environmentally) lower by some substantial degree. Thus equating on SES equates too much. The method would be a proper control of environmental factors if all children had been placed in their SES categories completely at random, in the nature of a true experiment. But as it is, SES classification is more a result than a cause of IQ variance. Consider the fact that there is a much lower

***Ed. Note: Footnote to ref. 69 was not received in entirety for inclusion in this print. Balance of footnotes at end of text..**

correlation between IQ and the SES in which one is reared than between IQ and persons' SES as adults. If SES *per se* were an important environmental determinant of IQ, we should expect children's IQs to correlate at least as much with the SES of their parents as with the SES the children attain as adults, but this is far from being the case. Burt (1961) found in England, with its high degree of socioeconomic stratification, that approximately 30 percent of the population changes SES (half going up and half going down) in each generation (based on father's occupation divided into six classes, from "higher professional" to "unskilled labor"). There is probably much greater intergenerational mobility in the United States, at least in the white population. Burt (1961b) found a correlation of .77 between fathers' IQs and their occupations (classified on a six-point scale) but a correlation of only .36 between children's IQs and their fathers' occupations. The standard deviation of parental IQs *within* occupational classes was 9.6; it was 14 for the children within occupational classes, only 1 point less than the *SD* of 15 for the total population. (The mean IQs of fathers and their children across the six occupational categories are shown in Figure 11.) This very great variance of children's IQs within each class is embarrassing to environmental theories. As Gottesman (1968, p. 37) has noted, ". . . the intelligent offspring of the dull parents and the dull children of the bright parents are phenomena difficult to account for on an environmental hypothesis . . . . Such findings are, however, completely predictable from the polygenic theory of intelligence."

Although matching for SES in comparing racial groups most likely works *against* a genetic hypothesis of the racial difference, because it matches

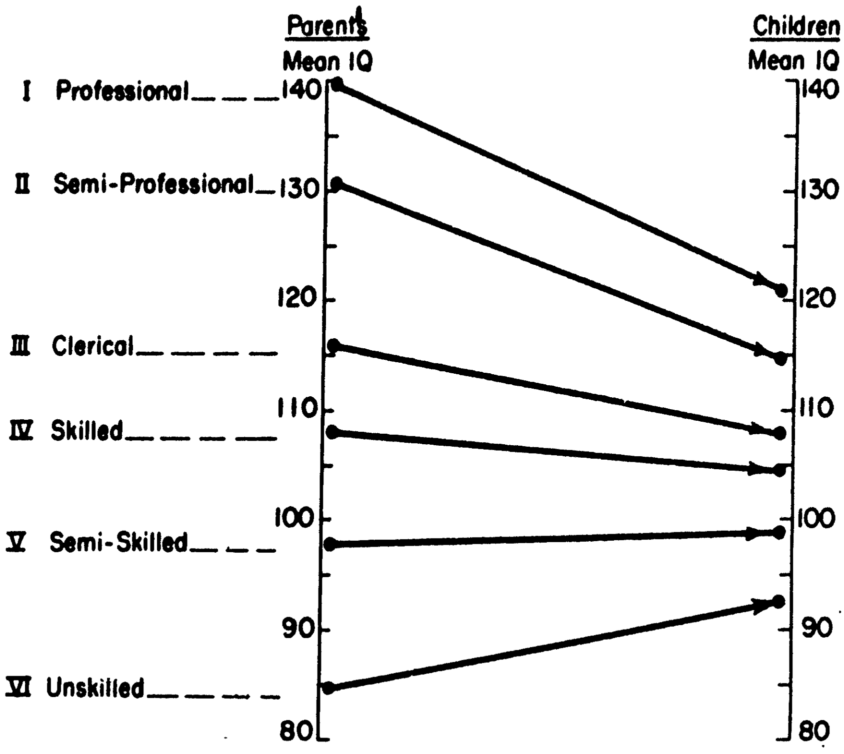


Figure 11. An illustration of the phenomenon of filial regression, originally established by Sir Francis Galton. (data from Burt, 1961)

to some degree for genetic as well as environmental factors, it is nevertheless instructive to note the results of studies which have attempted to control for SES by actual matching or by statistical equating of groups. In reviewing all the studies of this type up to 1965, Shuey (1966, p. 518) summarizes the results as follows: "With two exceptions, the colored averaged below the white groups in mental test performance in all of the 42 investigations. (The two exceptions were studies which showed ambiguous results or presented insufficient statistical analysis to permit an evaluation.) Average IQs were reported in 33 of the studies including a total of about 7,900 colored and 9,300 white Ss, and from these a mean difference of 11 points favoring the whites was obtained" (in contrast to a mean difference of 15-16 IQ points random samples are compared). "Twenty-five of the 41 studies were located in the North, and in at least fourteen of the researches the colored and white children were not only attending the same school, but were living in the same district or neighborhood. The combined mean difference in IQ between the 2,760 colored subjects tested in the North and the whites of comparable socioeconomic status or occupation was 7.6. Nearly all of these Ss in the eighteen studies were of school age, the whites and Negroes attending the same school and living in the same areas, many with large Negro populations."

A more recent study by Tulkin (1968) controlled not only SES but a number of subtle family environmental factors. Controlling SES alone did not overcome the racial difference in mean IQ. After the familial behavioral differences were equated, however, Tulkin concluded, "When family differences were also statistically controlled, there were no significant racial differences on test scores in the upper socioeconomic group, although differences remained significant in the lower socioeconomic group." Two critical points

should be made about this particular study, however. First, the upper-SES Negro group was small (*N* = 52), and though it did not show a statistically significant difference from the white upper SES group, the difference was in the same direction as in most other studies.[71] Second, Tulkin's analysis, which controlled (by covariance analysis) for various family factors within SES groups, was based on a composite score of Verbal and Nonverbal IQ plus five scholastic achievement tests. The composite score is thus more heavily weighted with scholastic achievement than with intelligence. As was noted previously, achievement scores (*a*) have lower heritability than IQ, (*b*) are more susceptible to family environmental influences, and (*c*) generally show smaller racial differences than does IQ, as is also true in Tulkin's study when the social and family variables are not controlled. Controlling these variables, therefore, should make a greater impression on achievement than on IQ tests. Even among the several achievement tests the magnitude of the difference between the upper SES white and upper SES Negro groups is greater for the *less* culturally loaded subject matter. The upper SES white sample exceeds the Negro upper SES sample, for example, by only 0.18 *SD* on the Language achievement test but by 0.51 *SD* on the Arithmetic test, a highly significant difference. When we compare the upper and lower SES white samples on these two achievement tests, on the other hand, the reverse occurs: the SES difference is greater for Language than for Arithmetic. These results, then, are consistent with the general finding, which is reviewed in a subsequent section, that the largest differences between Negroes and whites appear on tests that are the least culturally loaded. Tulkin's study also shows the composite achievement score to be more highly correlated with Verbal IQ than with nonverbal IQ, and the overall Negro-white difference is greater on nonverbal IQ. Tulkin's investigation might have

been more interesting if he had also applied the covariance control of family variables to nonverbal IQ alone rather than only to a composite score heavily weighted with achievement tests.

But as was pointed out, the method of matching racial groups for SES or other environmental variables and then comparing their mean IQs cannot tell us anything of importance, except that the SES matched groups are usually more alike in IQ than unmatched groups, for some indeterminate combination of genetic and environmental causes. We can go a step further, however, and seek a set of circumstances in which environmentalist and genetic theories should predict opposite results. The environmentalists' emphasis on equating for SES, and even for parental intelligence, is based on the idea that the SES variable has a predominantly causal connection with IQ, and therefore racial IQ differences will be eliminated to the extent that we are successful in equating SES and other environmental factors. The logic, at least, is clear, even if the premises are questionable. But the logic suggests an interesting comparison. What if we compared (a) Negro children reared in upper-middle class homes by Negro parents whose educational and occupational status and income were well above the average of the white population with (b) white children reared in the lowest SES category, whose parents are well below the average in intelligence, have less than an average education, and are either in unskilled work or on welfare? Even if parental IQs were not measured, there would be little doubt in such a case that the high SES Negro parents would have higher IQs in general than the low SES white parents. If these SES factors are more important determinants of IQ than genetic factors, there can be no doubt that the predicted result should be a much higher mean IQ for the upper SES Negro children than for the lower SES white children. An ideal study along these lines has not yet been done; it would involve obtaining

IQs of both parents of every child and making a prediction of the child's IQ based on a genetic model. We have already seen in Burt's data, shown in Figure 11, that there is some regression toward the population mean from parent's IQ to child's IQ. A genetic theory of the racial intelligence difference would predict, therefore, that Negro and white children would regress toward different population means. In the two SES groups we are considering here, the regression would be in opposite directions: the children of the high SES Negro parents would on the average regress to some degree downward toward the Negro population mean, and the low SES white children would regress upward toward the white population mean. Because the Negro population mean is about one standard deviation below the white population mean, the mean IQs of our two hypothetical groups of children would be much closer together than if we compared the mean IQs of low and high SES white children, and this should be so, according to our genetic hypothesis, even if the high SES Negro and white parents were perfectly matched on IQ. A corollary of this hypothesis is that Negro-white differences (based on children's IQs) should be greater for high SES than for low SES groups, since the IQs of the high SES Negro parents are presumably further from their population mean than the IQs of high SES white parents, making for a greater absolute regression effect in the Negro children. If the low SES Negro group is closer to the Negro mean IQ than is the high SES group, there will be less absolute <u>upward</u> regression of low SES Negro children's IQs than <u>downward</u> regression of the high SES Negro children.

Facts relevant to this hypothesis have been summarized by Shuey (1966, pp. 519-520): "Where Negro pupils have been compared with whites of the same occupational or socioeconomic class and where children from two or more classes have served as subjects, a greater difference has been found between

the racial samples at the upper than at the lower level." The eight relevant studies were all in agreement in this finding. Shuey (p. 519) continues: "The combined mean difference in IQ between the 617 colored *Ss* of higher status and their white counterparts is 20.3, in contrast with a combined mean difference of 12.2 between the 3374 colored and 2293 white children of low status." Overall, the mean IQ of the high status Negro children is 2.6 points *below* the mean IQ of the low status whites.[72] Since the publication of Shuey's review in 1966, this finding has been repeated in three major studies based on very large samples (Wilson, 1967; Coleman *et al.*, 1966; Scarr, 1971). In each study, when Negro and white children are classified by the same criteria into from 3 to 5 categories according to parental SES, the mean mental test scores of the lowest SES white group exceeds the mean IQ of the highest SES Negro group. No major study has found contradictory results. Data from the Coleman Report indicate that, with the exception of Puerto Ricans, other minority groups (American Indians, Mexican-Americans and Orientals), which are socioeconomically less advantaged than the white majority population, do not show this phenomenon -- that is, their upper SES group in every case exceeds the white lower (and usually also middle) SES group in test scores.

Other Minorities and Accentuated Environmental Inequalities

Negro-white IQ comparisons usually mean comparison of an environmentally less favored group with a more favored group. When the IQ difference is in the same direction as the environmental difference, the interpretation is problematic. Gottesman (1968, p. 34) has expressed the commonly held view that, "It is only when two individuals or two groups come from equally favorable environments that a difference in measured IQ can be interpreted to indicate a difference in genetic potential." But what about environmental

inequalities that are opposite in direction to the IQ difference? Kuttner (1968, p. 147) first noted the methodological possibilities suggested by this set of conditions: "If two populations can be studied which have experienced long-standing differential treatment, and yet both achieve at the same level, then grounds exist for presuming superior potential in one group. Or if one of the two groups responds to ameliorate conditions with a more markedly improved performance, then the same conclusion can be entertained. This procedure avoids artificially equating a disadvantaged group with a favored majority who may enjoy psychological and cultural benefits that are secondary products of status and hence beyond tabulation. At the same time, comparing deprived groups may isolate the significant variables that contribute most heavily to overall performance." Kuttner then proceeds with a detailed comparative analysis of Negro and American Indian environmental conditions and their mental and scholastic test performance.

On all the socioeconomic, educational, and health factors which sociologists have generally pointed to as causes of the Negro-white differences in IQ and scholastic achievement, the American Indian population has been about as far below Negro standards as the Negro ranks below whites. In 1960 Indian median income was 59 percent of Negro, which was 55 percent of white. Life expectancy, reflecting nutrition and health care, is much lower for Indians than for Negroes. In educational disadvantages, unemployment, poor housing, and infant mortality Indians are considerably worse off than Negroes. The Coleman Report (1966) used a scale composed of 12 categories of environmental variables[73] deemed important by social scientists as having a causal relationship to children's intellectual development. In this nationwide survey, which included more than 645,000 children in

4,000 public schools, Indians were lower than Negroes in all 12 environmental categories, and, overall, Indians averaged further below Negroes than Negroes averaged below whites. The relevance of these environmental indices is shown by the fact that _within_ each ethnic group they correlate in the expected direction with tests of intelligence and scholastic achievement. Since health, parental education, employment and family income, in addition to the 12 more subtle environmental factors rated in the Coleman study, are all deemed important for children's scholastic success, the stark deprivation of the Indian minority even by Negro standards, ought to be reflected in a comparison of the intelligence and achievement test performance of Indians and Negroes. The interesting fact is, however, that on all tests, from first to twelfth grade, Indians scored _higher_ than Negroes. Since many Indian children are bilingual, they can be most fairly compared with white and Negro children on nonverbal tests of intelligence, especially in the early school years. Coleman _et al_. (1966, p. 20) found that on a nonverbal intelligence test the mean score of Indian children in the first grade (approximately 6 years of age) exceeded the mean score of Negro children by 0.96 standard deviations, which is equivalent to about 14 IQ points. The first-grade intelligence test scores (with an overall national mean of 50 and a standard deviation of 10) of all the ethnic groups in the Coleman study (Table 9, p. 20) are shown below:

Group	Nonverbal	Verbal
White	54.1	53.2
Negro	43.4	45.4
Indian	53.0	47.8
Puerto Rican	45.8	44.9
Mexican-American	50.1	46.5
Orientals	56.6	51.6

77-843 O - 72 - 19

Thus, the Indian-Negro difference in a host of environmental factors is in just the <u>opposite</u> direction to the differences in mean performance on tests of nonverbal and verbal intelligence, reading comprehension, and math achievement.

Attempts to explain away these striking findings of the Coleman Report have invoked the ideas of unrepresentative sampling of the Indian population, effects of the racial composition of the school, and differences in motivation, self-concept, and educational aspiration between Negroes and Indians. For example, Bodmer and Cavalli-Sforza (1970, p. 27) write: "According to the Coleman report, however, American Indians typically go to schools where whites are in the majority, which is not the case for most of the schools attended by black children." Several comments about this statement are in order. It was pointed out earlier that Negro children in this study are about 1 <u>SD</u> below Indian children on the nonverbal test in the first grade. Since racial composition of the school <u>per se</u> has not been shown by the Coleman study or any other study to be related to achievement, it is most unlikely that the effect of racial composition of the school will have had sufficient effect by first grade to account for one <u>SD</u> IQ difference. Moreover, Coleman <u>et al</u>. (1966, p. 40) report that 48 percent of the Indians in the first grade sample were in schools in which the <u>majority</u> of pupils were Indians. If this argument of Bodmer and Cavalli-Sforza carried any conviction, we should predict that in the case where Negroes attend schools which have a majority of white pupils, they should do as well as Indians in similar circumstances. The Coleman report provides the conditions for examining this hypothesis (p. 40 and p. 243). At the twelfth grade, 92 percent of non-metropolitan North and West Negroes attend schools in which Negroes are in a minority; 91 percent of all 12th grade Indians attend schools in

which they are in a minority. Yet on the non-verbal intelligence test, non-metropolitan North and West Negroes score approximately 0.8 SD (equivalent to 11-12 IQ points) below the national average, while the Indians score about 0.1 SD (2-3 IQ points) below the national average. Even on the Verbal ability test the largely bilingual Indians exceed this Negro group by 0.4 SD (about 6 IQ points). Thus these data lend no support to Bodmer's and Cavalli-Sforza's conjecture. But they go on to argue that Coleman's Indians sample may not adequately represent the 70 to 80 percent of American Indians who live on reservations. This is mere surmise, but in any case it is irrelevant to the point being made by these data: it is the very same Indians who were tested in the Coleman study who also rated much lower than Negroes on all the environmental indices. Despite this environmental disadvantage, these same Indians scored higher than Negroes on the ability and achievement tests.

But what about motivation, self-esteem, and educational aspirations? These factors are commonly mentioned as explanatory variables in discussions of Negroes' mental test and scholastic performance. Gordon (1970, p. 254), for example, states: "Moreover, socially disadvantaged children have been determined by several investigators to be less highly motivated and to have lower aspiration for academic and vocational achievement than do their middle and upper class school peers." Further on in the same passage, Gordon (p. 255) writes: "As important as these attitudes toward school and learning may be, it is in the area of attitude toward self and others that the crucial determinants of achievement and upward mobility may lie..." Coleman et al. attempted to take account of these motivational and attitudinal factors. If poor environmental conditions, discrimination, and minority status depress academic motivation, aspiration, and self-esteem as a student, we should

expect the Indian students to show lower ratings on these variables than Negroes. If it is argued that Negroes suffer greater prejudice, discrimination, and the psychological handicaps they may engender, than do Indians, we should expect this to show up in Coleman's motivational and attitudinal assessments. But on a questionnaire of 16 items intended to assess school-related attitudes, motivation, self-concept, and educational aspirations, Negroes showed higher (more favorable) scores than Indians; in the twelfth grade, Negroes were higher on 14 of the 16 items. Overall, the several ethnic groups ranked as follows on these 16 motivational indices, from highest to lowest: whites, Orientals, Negroes, American Indians, Mexican-Americans, and Puerto Ricans. The three ethnic groups showing a lower standing than Negroes on the motivational measures all score higher on all tests of ability and scholastic achievement given in the first grade as well as in the twelfth grade, with only one exception -- 1st grade Puerto Ricans scored 0.05 SD (less than 1 IQ point) below Negroes on the Verbal ability test. Although these motivational indices correlate significantly in the expected direction with test performance within each of the ethnic groups (Coleman et al., 1966, p. 299, Table 3·221·1) showing that they are indeed relevant to academic attainment, the ordering of the several ethnic groups' mean test scores clearly do not correspond to their ordering on the motivational factors. At 12th grade, the rank-order correlation between mean test scores (the average of 5 ability and achievement tests) of the six ethnic groups and the rank order of their motviation indexes is 0.66. If the Negro group is omitted from the ranking, the rank order correlation becomes 1.0, i.e., perfect. The large rank order discrepancies on both sets of variables between the Negro and Indian groups can hardly be attributable to differential school drop-out rates; even at ages 16 and 17 Negroes have only a 3.1 percent higher enrollment rate than Indians (Coleman et al., 1966, p.450).

Even if all of the 3.1 excess of Indian dropouts consisted of the 3.1 percent with the lowest IQs in the Indian distribution, their not being included would raise the Indian mean test score by only 0.07 SDs (about 1 IQ point).[75] But the overall Indian-Negro test difference is 0.4 SDs (6 IQ points) at 12th grade, so at the very most only one-sixth of this difference could be attributed to differential dropout rates.

Coleman (p. 219) notes that Negro-white differences are more uniform across various tests than is true for all the other ethnic groups in the study: "The disadvantage for the various groups differs for different areas of achievement. From those from different linguistic cultures, Oreintal Americans, Mexican Americans, Indians, and Puerto Ricans, the disadvantage appears to be about the same for all areas tested....The Negroes' averages tend to be about one standard deviation below those of whites, which means that about 85 percent of Negro scores are below the white average."

Comparisons of white, Negro, and Mexican children in a California school district yield similar conclusions (Jensen, 1971*g*). Table 2 shows the results of this study in terms of the sigma units (*i.e.*, standard deviation of the test scores in the white sample) by which the minority group falls below the white group. The Stanford Achievement battery

consists of Word Meaning, Paragraph Meaning (Reading Comprehension), Spelling, Language (grammar), Arithmetic Computation (mechanical arithmetic), Arithmetic Concepts, and Arithmetic Applications (thought problems). The nonverbal intelligence tests were a composite of the Lorge-Throndike non-verbal IQ, Raven's Progressive Matrices, and Gesell's Figure Copying Test (see Figure 1). The Home Index (Gough, 1949, 1971) is a 25-item inventory

Table 2

Number of White Sigma Units by which Minority Group Means Fall Below the White Mean

	Sample Size (N)			Stanford Achievement Tests		Nonverbal Intelligence		Home Index (SES)		Adjusted Achievement Means	
Grade	White	Negro	Mexican	Negro	Mexican	Negro	Mexican	Negro	Mexican	Negro	Mexican
1	285	218	258	.25	.34	1.07	.53	—	—	-.09	.15
2	229	162	250	.57	.37	1.03	.70	—	—	.15	.06
3	281	207	241	.83	.68	0.98	.53	.58	1.13	.11	.05
4	237	189	239	.69	.59	0.95	.48	.38	1.18	.17	.15
5	242	198	211	.75	.54	1.05	.62	.70	1.18	.21	.10
6	219	169	218	.84	.69	1.23	.67	.47	1.36	.09	.02
7	388	262	305	.71	.57	1.13	.72	.71	1.36	.07	.08
8	356	289	303	.64	.62	1.18	.79	.77	1.34	.06	.08
Mean				.66	.55	1.08	.63	.60	1.26	.10	.09

of socioeconimic status based on educational and occupational level of parents, material possessions in the home, parental participation in the middle-class and upper-middle class social and civic activities, and cultural advantages in the home, *e.g.*, music lessons and art. It can be seen in Table 2 that at every grade level from 1 to 8, the Negro group is further below the white group than is the Mexican group, and the difference is greater for the nonverbal tests than for the scholastic achievement tests. Yet on the Home Index, the Mexicans are further below the Negroes than the Negroes are below the whites. The relevance of the Home Index is shown by its positive correlations with test performance within groups, and in a multiple-regression equation for predicting scholastic achievement the Home Index makes a unique contribution to the overall prediction of achievement. Also a questionnaire similar to that used in the Coleman study to reflect attitudes of self-confidence, self-esteem, and educational aspirations showed only small Negro-white differences, while scores were generally much lower for the Mexican group. None of these indices reflects the added disadvantage of the Mexicans' bilingualism. In the present sample, the percentage of Mexican children whose parents speak only English at home is 19.7 percent as compared with 96.5 percent for whites and 98.2 percent for Negroes. In 14.2 percent of the Mexican homes Spanish or other foreign language is spoken exclusively, as compared with 1.1 percent for whites and 0.5 percent for Negroes. Many of the parents of the Mexican children grew up in Mexico where they had little or no education. Most of them came to the central valley of California (in which the present study was conducted) as agricultrual workers living in overcrowded, unsanitary migratory camps that follow the fruit and vegetable crops. Because of the nomadic life of the parents, many of these children have poor records of school attendance. The adjusted Achievement means (the last two columns of Table 2) refer to

the achievement test means after they have been adjusted by analysis of covariance using intelligence and SES as the control variables. In effect, these two columns represent the sigma units by which the minority groups fall below the white in achievement when all groups are statistically equated for intelligence and SES. The achievement differences that remain are practically negligible and can be even further reduced by including additional control variables, such as motivational and personality tests, in the covariance analysis.

On Raven's Progressive Matrices, a nonverbal, culture-fair test of the *g* factor of intelligence, the Mexicans were intermediate between whites and Negroes, as shown in Figure 12, despite the much lower SES and poorer motivation of the Mexican pupils.

Finally, a factor analysis was performed on the intercorrelations among all the variables in all three ethnic groups combined. Four major factors emerged: (I) Scholastic Achievement and Verbal Intelligence, (II) Nonverbal Intelligence, (III) Rate Memory Ability, (IV) Socioeconomic Status. Minor factors were (1) speed,motivation, and persistence, (2) Neuroticism, (3) Extraversion. (4) Age in months. These variables are, in effect, partialled out of the major factors. Since the four major factors are orthogonal (i.e., uncorrelated with one another) by virtue of the type of factor analysis used (varimax rotation of the principal components), each one can be viewed as a "pure" measure of a particular factor in the sense that the influences of all the other factors are held constant. Factor scores were obtained for every pupil on each of the four main factors. (The factor scores have an overall mean of 50 and a standard deviation of 10.) The mean factor

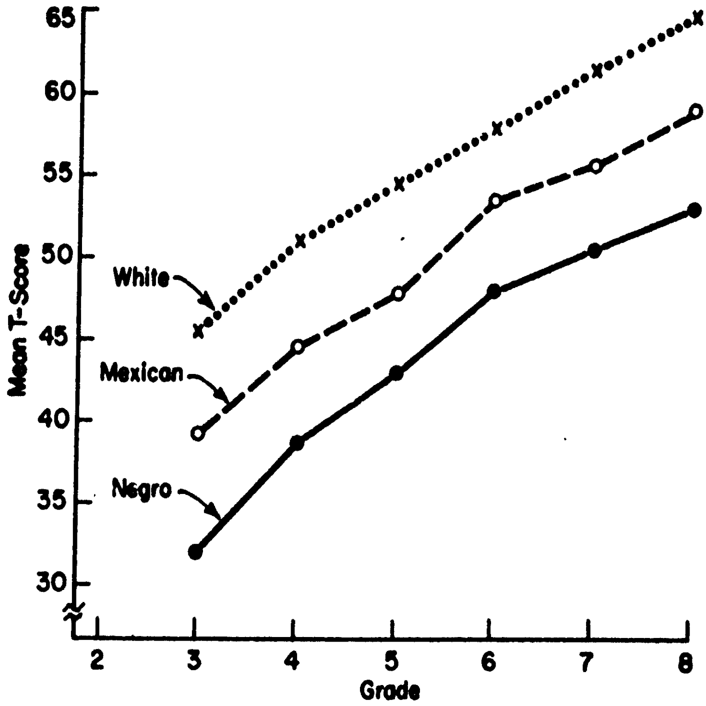


Figure 12. Mean T scores ($\bar{x}$ = 50, SD = 10) on Raven's Progressive Matrices.

scores of the three ethnic groups are shown for Grades 4, 5, and 6 (total N = 1,179) in Figure 13. On Factor I (Verbal IQ and Achievement) all three

ethnic groups differ significantly from one another. On Factor II (Nonverbal IQ) the Negro-white and Negro-Mexican differences are significant, but the Mexican-white difference is not. On Factor III (Rote Memory) the only significant difference is between Mexicans and Negroes at grades 4 and 5. On Factor IV (SES) the Mexicans fall significantly below whites and Negroes, whose SES factor scores differ only slightly in this school population.

The test results for various minorities reported in the Coleman study in many ways are paralleled by the percentages of the various groups employed in professions that depend upon educational attainments. Weyl (1969) has made these comparisons, based on the 1960 U.S. Census, in terms of an index consisting of the raito of the total proportion of the ethnic minority in the profession to the statistical expectation, which is the proportion of the total population constituted by the ethnic minority times the proportion of the population constituted by the members of the profession. An index value of 100 means the ethnic group is represented in a given profession according to statistical expectation; one of 50 means that it contributes half the expected number of professionals, and one of 200 means it supplies twice the statistical expectation. The index figures are shown in Table 3.

It is interesting to note that Orientals, who in Coleman's study scored higher than any other groups in non-verbal and mathematical abilities, have the highest index figures in accounting, architecture, engineering, and natural sciences. Negroes, who were lowest in nonverbal abilities and relatively

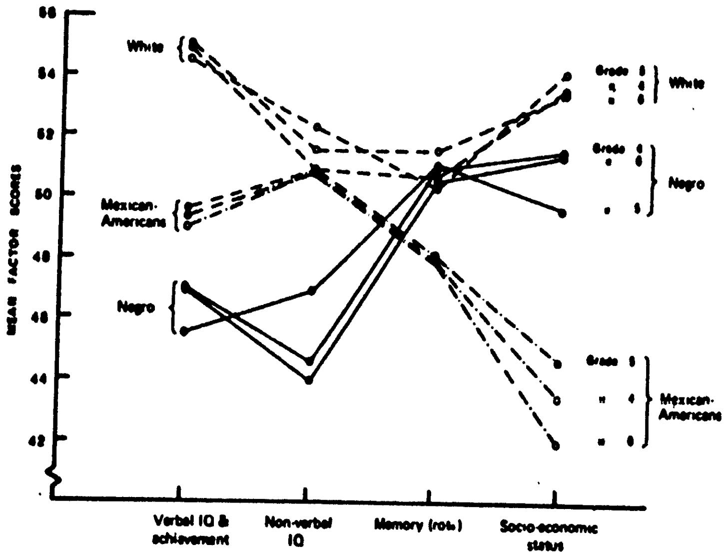


Figure 13. Factor scores (mean = 50, SD = 10 within each grade level) for four variables, comparing white, Mexican-American, and Negro samples in grades 4, 5, and 6. The factor scores are independent of one another; that is, the scores on any one factor reveal differences between subjects who are statistically equated on the three other factors. (From Jensen, 1971a, Table 6)

Table 3

Index Figures of the Contribution of Five Ethnic Groups
to American Professions in 1960*

Profession	White	Negro	Indian	Japanese	Chinese
Accountants	112	7	38	166	174
Architects	110	5	0	232	506
Artists and Writers	110	16	133	209	136
College Professors	107	32	0	143	537
School Teachers	103	76	86	120	318
Engineers	111	5	57	124	303
Natural Scientists	109	20	0	205	438
Lawyers and Judges	111	11	19	54	53
Clergyman	104	66	124	89	23
Physicians	108	21	10	182	302
Nurses	106	54	124	116	76
Technicians	107	36	86	201	197
All	107	38	76	139	189

* From Weyl (1969, p. 114)

higher in verbal, show the lowest indexes for professions involving spatial and quantitative abilities, such as architecture and engineering, and are most heavily represented in such verbal professions as school teaching and the clergy.

If prejudice and discrimination are more important than abilities in determining a group's representation among the professional classes, then it should be puzzling that two minorities -- the Japanese and Chinese -- who have also been subject to discrimination and other social disadvantages in the United States should have considerably higher indexes than the white majority. The group labeled white in Table 3 includes Jews. Their separate overall average is an index of 282, which is by far the highest, and nearly triple the index for non-Jewish whites, although Jews have been subject to prejudice and social discrimination. The figures of Table 3 lend some support to the popular characterization of Jews and Orientals as America's intellectual elite. The reasons, undoubtedly complex, probably involve selective migration, selective and assortative mating patterns, and other associated genetic and cultural factors.

Inequality of Schooling

Some writers have pointed to supposed educational inequalities as a cause of poor Negro performance in IQ and achievement tests. Thus, Bodmer and Cavalli-Sforza (1970, p. 27) write: "Black schools are well known to be generally less adequate than white schools, so that equal number of years of schooling certainly do not mean equal educational attainment." This statement clearly implies that the Negro-white scholastic achievement gap (generally equivalent to 2 to 4 grade levels at high school graduation) is attributable, at least in large part, to the superior school facilities enjoyed by white children. But there is now massive evidence which clearly contradicts this claim.

The well-known Coleman report (Coleman _et al_., 1966) was funded by the U.S. government specifically to determine the degree to which inequalities in educational performance are attributable to inequalities in school facilities. This enormous survey of the nation's schools found that very little (overall, less than 10 percent) of the variance among schools in scholastic achievement was due to differences in school facilities, including variables such as physical facilities, class size, curricula, teacher salaries, experience and qualifications, special services, etc. The report concluded: "Differences in school facilities and curriculum which are the major variables by which attempts are made to improve schools, are so little related to differences in achievement levels of students that, with few excpetions, their effects fail to appear even in a survey of this magnitude" (p. 316). More specifically, the major findings of the Coleman study are summarized as follows (p. 325):

"Taking all these results together, one implication stands out above all: That schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. For equality of educational opportunity through the schools must imply a strong effect of schools that is independent of the child's immediate social environment, and that strong independent effect is not present in American schools."

In an analysis of relationships between (_a_) minority enrollment, (_b_) IQ, and (_c_) reading scores, on the one hand, and (_d_) pupil expenditure, (_e_) teacher salary, (f) pupil/teacher ratios, and (_g_) number of school administrators, on the other hand, in 191 school systems in California, it

was found that the school-related variables have negligible correlations with IQ and reading scores, while percentage of minority enrollment has very high negative correlations with the school's mean IQ and reading level. At the same time, there is a slight, but nonsignificant, positive correlation between minority enrollment and the indices of school quality (Jensen, 1971a).

The fact is that the achievement level in a school is predictable from a number of demographic characteristics over which the school itself has no control whatsoever. Thorndike (1951), for example, correlated average IQ and an average scholastic achievement index (based on half a million children) with 24 census variables for a wide range of communities, large and small, urban and rural. Eleven of the correlations were significant at the 1 percent level. Census variables showing the highest correlation with IQ and achievement were educational level of the adult population (.43), home ownership (.39), quality and cost of housing (.33), proportion of native-born whites (.28), rate of female employment (.26), and proportion of professional workers (.28). In a multiple correlation these census variables predict IQ and achievement between .55 and .60.

Statistics based on all schools (over 900) in New York City show a strong negative correlation between pupil expenditures and scholastic achievement, since the school's financial resources are positively correlated with the proportion of Negro and Puerto Rican enrollment (Gittell, 1971). The 30 elementary schools in New York with a per pupil expenditure of more than $1,100 per year (mean of $1,330) showed reading and arithmetic scores five to seven months below the scores of pupils in the 101 schools with an expenditure below $600 (mean of $551). Teacher-pupil ratios in the high-scoring schools were more than twice as high as in the low-scoring schools. In

other words, by any objective indices of advantages provided by the majority of schools in New York, the minority children are more favored than majority children. The report states:

"The evidence we have accumulated is somewhat surprising. We have recorded traditional variables that supposedly affect the quality of learning: class size, school expenditure, pupil/teacher ratio, condition of building, teacher experience, and the like. Yet, there seems to be no direct relationship between these school measurements and performance. Schools that have exceptionally small class registers, staffed with experienced teachers, spend more money per pupil, and possess modern facilities do not reflect exceptional academic competence" (Gittell, 1971, p. 2)

Jensen (1971a) compared large representative samples of Negro and Mexican-American pupils with white pupils from kindergarten through 8th grade in largely de facto segregated schools in the same California school district, using a comprehensive battery of tests of mental abilities and of scholastic achievement, in addition to personality inventories and indices of socioeconomic and cultural disadvantage. It was found that when certain ability and background factors over which the schools have little or no influence are statistically controlled, there are no appreciable differences between the scholastic achievements of minority and majority pupils. The study lends no support to the hypothesis that the schools are discriminating unfavorably against Negro pupils, whose average scholastic achievement was .66 SD below the white mean. (On nonverbal tests of intelligence, the average difference was 1.08 SD.) Furthermore, it was found that Negro children are as far below the white IQ mean, in sigma units, at kindergarten or first grade as at 12th grade. If the schools contributed to the Negro-white IQ difference, one should expect to find an increase in the difference from kindergarten to 12th grade. When race is entered as a variable

into a multiple regression equation, along with a number of measures of mental ability and social background, to predict scholastic achievement, race *per se* makes no significant independent contribution to the prediction. This means, in effect, that a Negro pupil and a white pupil who are matched for IQ and home background will perform equally well in school. The average Negro-white achievement difference is thus related to race only incidentally, through association with intelligence. There is no evidence that the schools have contributed to this difference, and, in fact, there is evidence to the contrary -- schools tend to have a leveling influence, so that Negroes and whites actually differ less in scholastic performance than in intelligence as measured by nonverbal tests at an age before the school could have had any appreciable impact.

The fact that Negro children have been shown to lag less far behind white children on scholastic achievement tests than on non-verbal intelligence tests which tap skills that are not taught in school (but which predict scholastic performance) belies the theory which blames the Negroes' lower achievement on poor teaching by the schools. Kenneth Clark (1968), for example, rejects the cultural deprivation theories of Negro scholastic performance:

"The picture of deprivation given by these theories is one of total stark, bleak deprivation. The degree of poverty in urban working-class Negro homes is so stark that the child has absolutely no sensory stimulation whatsoever. . . many of these studies talk about lower class culture as if it were totally isolated from all communication with the rest of our society. Not one of these reports, to my knowledge or memory, ever talks about the reality: there is no sub-culture in our large society that is so deprived as to be unable to have some communication with the larger culture through our mass media, television, etc. . . The sophisticated version of the cul-

77-943 O - 72 - 20

tural deprivation explanation of academic retardation for Negro children has seemed to have built up a mythology of cultural isolation that does not seem to be supported by reality" (pp. 181-182). Clark then goes on to blame the schools for the academic lag of Negro children:

"To what extent are they not being taught because those who are in charge of teaching them do not believe that they can learn, do not expect that they can learn, and do not relate to them in ways that are conducive to their learning?" (p. 183).

This is a currently popular hypothesis, but I can find no objective evidence that supports it. It is not sufficient merely to note that there are some teachers and schools which have undesirable attitudes toward minority children and provide an inferior educational environment for such children. We must examine those schools which have taken pains to give Negro children every advantage, and even more, than is provided for the white children, and in which the teachers, both white and Negro, have been specially selected for their dedication and favorable attitudes toward minority pupils. Where these conditions exist, has there been found an appreciably smaller achievement gap between the races? I believe there has been some reduction in the achievement gap in the California schools in which I have collected data and where there has been a concerted effort to give every advantage to Negro children. It is under these conditions that the scholastic achievement difference is about 30 percent less than the average difference in intelligence, when intelligence is measured by very non-scholastic, nonverbal tests. There is no reason to believe that good teaching and good educational facilities will not improve Negro scholastic performance. But there is equally no evidence to support the belief that the Negro-white difference that still persists under these conditions is a result of some subtle, invisible discrimination by teachers whose

attitudes and expectations depress Negro performance.

Teacher Expectancy

Credence in the notion that lower Negro performance on IQ tests results from teachers' expectations was widely promulgated by one of the most highly publicized studies in the recent history of educational research--the famous Pygmalion in the Classroom study by Rosenthal and Jacobson (1968). The main thesis of these authors is that a teacher's expectation of what pupils are able to do creates a "self-fulling prophecy" which actually raises or lowers the children's IQs and level of scholastic achievements. Thus, initial differences in test scores, if known to the teachers, should become magnified in subsequent testings as a result of teacher expectations. And similarly, on the basis of previous experience, preconceptions, etc. concerning the relative abilities of Negro and white children, teachers' expectations should, according to this hypothesis, create or magnify performance differences between Negro and white pupils.

Rosenthal and Jacobson (1968) attempted to test this hypothesis by having teachers administer a group paper-and-pencil intelligence test to all pupils from kindergarten through sixth grade in a South San Francisco elementary school. Teachers were told that the test was intended ". . . to predict which youngsters are most likely to show an academic spurt." In September, each teacher was given a list of those children (actually selected by a table of random numbers) who were supposedly predicted by the test to be most likely to show an academic spurt during the school year. The children were tested again by the teachers in January and May. The authors' conclusion, which has been repeated and accepted so widely in educational circles, is that the teachers' expectancies influenced the mental development (or test performance) of the children.

But the evidence presented in the study does not in the least support

this conclusion, as is emphatically pointed out in the major critical reviews of the study (Snow, 1969: Thorndike, 1968). In the first place, the data themselves present so many bizarre features as to make them totally suspect. For example, in one grade the control group (i.e., non-expectancy of a spurt) had a mean IQ of 31! This is just barely at the imbecile level; such defective children actually are never enrolled in regular classes. Even if we accept the authors' conclusions without questioning the quality of the data or their analyses, the "prophecy" effect shows up in only 19 pupils in two grades (one of which has the control group with a mean IQ of 31). As Thorndike comments in his review: "If these present data show anything, they show that the testing was utterly worthless and meaningless" (p. 710). Thorndike concludes that the study ". . . is so defective technically that one can only regret that it ever got beyond the eyes of the original investigators! Though the volume may be an effective addition to educational propagandizing, it does nothing to raise the standards of educational research . . . In conclusion, then, the indications are that the basic data upon which this structure has been raised are so untrustworthy that any conclusions based upon them must be suspect. The conclusions may be correct, but if so it must be considered a fortunate coincidence."

But <u>are</u> the conclusions of Rosenthal and Jacobson, in fact, correct? Fortunately, the expectancy hypothesis has since been subjected to rigorous tests with the proper controls and appropriate methodology. Since Rosenthal and Jacobson reported finding the strongest expectancy effect in the first grade pupils, Claiborn (1969) attempted to demonstrate the effect, using procedures similar to those of Rosenthal and Jacobson, in twelve first-grade classes. He found no evidence of the expectancy effect.

The largest study, by Fleming and Anttonen (1971), involved 1,087 second grade pupils in 39 classrooms in 22 schools representing two socio-economic levels -- low SES and middle SES. The design of this study was more complex than that employed by Rosenthal and Jacobson, so that the influences of a number of factors could be assessed -- the effect of the teachers' attitudes toward intelligence tests, the effect of giving the teacher the results of the tests vs. withholding test scores, the effect of giving the teacher grossly inaccurate IQs (inflated by 16 points) on some children, and the differential effect of all these variables on children's retest performance as a function of SES. Two intelligence tests were used (Kuhlman-Anderson and Primary Mental Abilities). Pre- and post-testing occurred at the beginning and end of the school year. All post-testing was conducted by graduate assistants who were unaware of the nature or purpose of the study. The results of the post-test analysis revealed no significant differences among the four treatment groups (viz., a. Teachers given IQ scores, b. Withholding of IQ information, c. Teachers given Primary Mental Abilities percentiles, d. Teachers given IQs inflated by 16 points). There was a significant effect of teacher's opinion of IQ tests as assessed by a questionnaire, but the effect appeared only for the middle SES children. When teachers were categorized into three groups (High, Middle, and Low) on their opinion of the validity of IQ tests, the Low opinion teachers' pupils, in the middle SES classes, received significantly lower IQs than were obtained by pupils whose teachers had a high opinion of IQ tests. The effect was in the same direction for low SES children, but was so small as to be non-significant even with the large sample sizes employed. When teachers were asked to assess the accuracy of the IQ scores given to them, based on knowledge they gained of the child throughout the school year, they significantly judged the IQs inflated

by 16 points as less accurate than the regular IQs. Fleming and Antonnen (1971, p. 250) conclude: "It appears that, in the real world of the teacher using IQ test information, the self-fulfilling prophecy does not operate as Rosenthal hypothesizes. We can only conclude that teachers are more sensitive to the functioning level of students than previously believed, since teachers, in fact, identified the inflated group as less accurate. Recognition of the deception by the teachers suggests that day to day living with the academic performance and behavior of children, at least for this group of teachers, provides more input than the results of an intelligence test administered on one given day."[76]

Various Motivational Factors

A number of motivational factors have been investigated in attempts to explain at least some of the Negro-white difference in intelligence test performance in terms of differences in motivation. The evidence to date does not support the differential motivation hypothesis. This should not be too surprising, since experimental studies of the effects of motivational factors on intelligence testing have generally shown either very small or nonsignificant effects, and when differences have been found they tend to show that conditions most typical of those in which intelligence tests are normally given yield the best scores. Burt and Williams (1962), for example, found that children obtained slightly higher scores when taking tests for school promotion rather than for experimental purposes. Intelligence tests are quite insensitive to external motivational manipulations. Tiber and Kennedy (1964) tested middle and lower-class white children and lower-class Negro children with and without several different incentives, such as praise after each test item, verbal reproof, and candy reward. These various testing conditions had no significant effects on Stanford-Binet IQs and

showed no interaction with social class or race. Tiber and Kennedy concluded that the IQ differences usually found between such social class and racial groups cannot be attributed to motivational differences. This conclusion is too sweeping, of course, since other motivational factors not under the experimenter's control could affect test performance. But the fact remains that scores on IQ tests have proved highly resistant to experimental manipulations of incentives and motivational sets.

Self-Concept

The testee's self-concept or self-esteem has been claimed to be an important determinant of test performance, and Negro-white IQ differences have been attributed to the purported lower self-confidence and self-esteem of Negro children, at least in test situations. The dozen or so studies of this topic are about evenly divided in supporting or failing to support some hypotheses related to this issue (Zirkel and Moses, 1971). But the present evidence is so ambiguous that no really strong conclusion in either direction seems to have emerged. What has not been at all consistently shown, however, is that assessments of self-concept (or self-esteem, etc.) jointly (a) differ for Negroes and whites, (b) are correlated with IQ or scholastic achievement, and (c) are not merely a reflection of the pupil's more or less objective appraisal of his own scholastic standing and aptitudes. Inventories intended to assess the pupil's self-concept of his abilities typically contain items such as, "I feel that I just cannot learn," "How do you compare in ability with your friends (or classmates)?" "Do you have the ability to complete high school (or college)?" etc. (e.g., Anderson & Johnson, 1971, p. 295). If pupils' answers to such questions in any way reflect an awareness of their standing among their age mates in scholastic ability, it should not be surprising that their self-concept scores are correlated with objective measures of intelligence and scholastic perfor-

mance. Children are quite perceptive about the relative standing of themselves and their schoolmates in "brightness."

Coleman et al. (1966, p. 323), in the largest study employing a pupil attitude and self-concept inventory, found one attitude questionnaire item which, far more than any others, differentiated minority and white children and also correlated with scores on intelligence tests. It was the item "good luck is more important than hard work for success," which is referred to as a "control of environment" attitude. The largest differences, which were found in the ninth grade, indicate that those minority pupils (Negro, Mexican American, Indian, and Puerto Rican) who disagree that "good luck is more important than hard work" obtain significantly higher verbal test scores, on the average, than white pupils who agree with the statement. Thus it is clear that the factor tapped by this particular question is correlated with verbal intelligence (to about the same degree in all ethnic groups) and also shows significant ethnic group differences. (The correlation of this item with nonverbal intelligence is considerably lower, but is still significant.) But the causal connection between response to this questionnaire item and intelligence scores is not established. Does the attitude directly affect test performance, or are less intelligent pupils merely more likely to attribute success to "good luck" rather than to "hard work"? The latter explanation seems more probable. Gough (1953) was able to produce a noncognitive "intelligence test," made up of "personality"-type questions, which correlated remarkably with scores on standard intelligence tests. Not a single item of the Gough questionnaire calls for mental ability per se. All the items are questions such as "I have often been frightened in the middle of the night" (keyed False), and "I gossip a little at times" (keyed True). Few would argue that being frightened

at night will lower one's intelligence, or that by gossiping one can raise his IQ. Belief in "luck" probably falls into this same category of attitude items that comprise Gough's non-intellectual intelligence test.

The study which has used what is probably the most elaborate and most reliable index of self-esteem, the 42-item Coopersmith Self-Esteem Inventory, administered to groups of white, Negro, and Puerto Rican 5th and 6th graders, matched for SES and IQ, came to this conclusion: "Support was thus given for the growing number of studies which indicate that the self-concept of Negro children does not differ significantly from and may even be higher than that of white children. It also appears that the self-concept of Puerto Rican children is significantly lower not only than the self-concept of white children, as shown in the minimal amount of previous research, but also than that of Negro children" (Zirkel & Moses, 1971, p. 260). The fact that the ethnic groups were selected so as to be highly similar in SES and IQ makes the results rather tenuous. If it is claimed that lower IQs are partly a result of poor self-concept, then matching ethnic groups for IQ and SES could well minimize differences in the self-concept scores.

Competition and Failure Threat. Still another motivational theory of low Negro IQ and scholastic attainment, originally suggested by the experimental research of Irwin Katz (1964, 1968), holds that Negro test performance is depressed by a constellation of factors comprised of (a) failure threat -- the Negro's expectancy of a low probability of success in competition with whites or white norms on an intelligence test, and (b) social threat -- emotional responses of fear, anger, and humiliation that are presumably detrimental to performance and may be elicited by a white examiner, especially

if the examiner is perceived as unsympathetic, supercilious, and authoritarian. Katz has tested these hypotheses experimentally by administering test-like tasks to Negro college students with and without instructions that it was or was not an intelligence test, that the testees were or were not competing with whites or white norms, with white or Negro examiners, threatening or friendly examiners, and with or without external threats such as strong electric shock. This research, although interesting and important in its own right, has unfortunately been misrepresented as indicating that these situational factors manipulated in Katz's experiments affect Negro performance on standard intelligence tests and in situations that are typical of those in which intelligence tests are ordinarily administered (e.g., Watson, 1970). It has not been demonstrated that the effects hypothesized by Katz account for any of the Negro-white difference in IQ as measured by any of the standard individual or group administered tests of intelligence. Whether or not it is possible significantly to influence subjects' performance on certain experimental tasks, specially selected for their sensitivity to distraction and emotional arousal, under conditions that are very atypical of ordinary intelligence testing (such as threatening instructors, examiners acting hostile and authoritarian, and threat of electric shock while performing the task) is not at issue.

The several experiments of Katz and his co-workers (recently reviewed by Sattler, 1970)[77] did not use intelligence tests, but timed experimental tasks depending mainly on speed of performance, rather than mental power. Such speed tasks are known to be more sensitive to distractions, emotional states, and the like. One of the tests, for example, was simple aritheatic -- but the subjects were college students. This makes sense in terms of the effect Katz was trying to detect in his experiment. The aim was to use a test which was so easy that not intelligence but mainly a speed factor,

highly sensitive to distraction, would be the greatest source of variance in the experiment. The experimental tasks that come closest to resembling anything found in standard intelligence tests are the digit-letter substitution and digit symbol tests, which resemble the digit symbol subtest of the Wechsler Intelligence Scale. But of the eleven subtests comprising the Wechsler, Digit Symbol has by far the lowest loading on *g* (correlated for attenuation) and the lowest correlation of any subtest with the total IQ. Thus the tests used by Katz could hardly have been a better selection if the aim was to reveal the effects of situational variables on performance. But they were not intelligence tests and the conditions of administration that produced lower scores were not typical of normal testing. Moreover, Katz used Negro college students, and since college students are selected mainly for intelligence, this would have the effect of narrowing the range of variance that intelligence might contribute to performance on the tests, permitting personality and emotional factors to contribute a relatively larger proportion of the variance. Then, too, it should be noted that the Katz experiments are not concerned with comparing Negro and white performance on tests, but with showing variation in Negro performance under different testing conditions. So we do not know how much Negro-white difference on any test would be accounted for by the Katz hypotheses. The magnitude of the score decrements found by Katz, even under the most extremely unfavorable conditions, are small in relation to the standard deviation in the population and do not invariably show up in the predicted direction from one experiment to another. When results are in the opposite direction to the hypothesis, it seems not to cast doubt on the hypothesis but to give rise to *ad hoc* rationalizations, such as, "In the last study the results when the tester was Negro were in the opposite direction, regardless of the kind of feedback used.

There may be a simple regional explanation for these contradictory findings, since the earlier experiment was done in Florida, and the latter one in Tennessee. Perhaps the Negro student in the Deep South is more fearful of competition with white peers than is the Negro student in the Upper South" (Katz, 1968, p. 281).

Race of Examiner. Despite the many conjectures (see Sattler, 1970, pp. 143-144) that the race of the examiner affects Negro-white differences on actual intelligence tests, the rather meagre evidence which exists does not support this belief. The most adequate study intended to examine the testee and tester racial interaction used three Negro and three white female testers giving the Stanford-Binet Intelligence Test and the Peabody Picture Vocabulary Test to Negro and white children enrolled in a Head Start program in Tennessee. Race of the examiners was found to have no significant effect on the test scores of either the Negro or the white children (Miller & Phillips, 1966). Shuey (1966) compared the 19 studies of Negro IQ in elementary school children in the South where the testing was done by a Negro with the results obtained on all Southern Negro School children. Shuey concludes: "The 2,360 elementary school children tested by Negroes earned a mean IQ of 80.9 as compared with a combined mean of 80.6 earned by more than 30,000 Southern Negro school children, an undetermined but probably a large number of whom were tested by white investigators. The present writer also calculated the combined mean IQ achieved by 1,796 Southern colored high school pupils who were tested by Negro adults. This was 82.9 as compared with a mean of 82.1 secured by nearly 9,000 Southern colored high school students, many of whom were examined by white researchers. From these comparisons it would seem that the intelligence score of a Negro school child or high school pupil has not been adversely affected by the presence of a white tester" (p. 507).

The most recent and comprehensive review of this topic concludes: "The experimenters' race affects subjects' picture and doll preferences, but may not influence their scores on intelligence tests and personality measures" (Sattler, 1970, p. 137).

Differential Test Performance. If generalized attitudes that depress Negro but not white performance accounted for the Negro-white IQ difference, it would be hard to explain why some kinds of tests are so affected and not others, for Negroes do not perform poorly on all kinds of tests. Jensen (1968b) has shown, for example, that Negro preschoolers with a mean IQ 19 points below white children perform equal to the whites on tests of memory span -- when the latter tests are given under the same conditions and by the same examiner as the IQ tests. But a factor analysis showed that the memory tests were not measures of intelligence; they involve another kind of mental ability. In this study, the memory test actually called for more attention and freedom from distraction than did the IQ test. Subsequent studies (Jensen, 1970b,c; 1971a; Jensen & Rohwer, 1970) have consistently found much smaller or nonsignificant Negro-white differences on tests of immediate memory while at the same time there were differences of more than one standard deviation on intelligence tests administered by the same testers under the same conditions as the memory tests. If motivational factors or testee and tester interactions affect the intelligence score, one would have to explain why these factors do not affect the memory test scores. It appears that, in general, the degree to which a test does not correlate with intelligence or abstract, conceptual, problem-solving ability, it fails to show a mean difference between Negroes and whites. This observation affords a means for assessing motivational differences in test performance more or less uncontaminated by differences involving intellectual ability *per se*. If it is

hypothesized that poor test performance results from poor motivation, inhibition of effort, or just not trying as hard as others, it is difficult to pit this hypothesis against one which states that the difference in test performance is due to a lesser cognitive ability for dealing with *g* material, *i.e.*, concepts, relationships, and abstractions, if one and the same test is used to assess motivation and intelligence, for then motivation and mental ability are confounded in the testing situation.

To get around this methodological problem, a motivation-sensitive test involving speed and persistence was devised so as to maximize dependence upon effort and minimize dependence on cognitive ability, particularly of the kind characterized by *g*. This experimental task, called the Making X's test, is one kind of objective assessment of test-taking motivation. It gives an indication of the subject's willingness to comply with instructions in a group testing situation and to mobilize effort in following these instructions for a brief period of time. The test involves no intellectual component, although it may involve some motor skills factor. But most of the individual differences in scores is attributable to the subjects' motivation and effort. The test consists of two parts. On Part I the subject is asked simply to make X's in a series of squares for a period of exactly 90 seconds. In this part the instructions make no mention of speed; they merely tell the child to make X's in the "boxes." The maximum possible score on Part I is 150, since there are 150 "boxes" provided in which the child can make X's. After a 2-minute rest pause, the child turns the page of the test booklet to Part II. Now the child is instructed to show how much better he can perform than he did on Part I and to work as rapidly as possible. Again, he is given 90 seconds to make as many X's as possible in the 150 "boxes" provided. The significant improvement in score from Part I to Part II shows that the test is sensitive to the motivating instructions.

This test has also shown greater sensitivity to teacher and experimenter differences than is found with intelligence or achievement tests. The Making X's test was given to all the 4th, 5th, and 6th graders in an urban school system (1588 whites and 1242 Negroes). At each grade level, the Negro mean score was equal to or slightly <u>higher</u> than the white, and the gain from Part I to Part II was significantly <u>higher</u> for Negroes than for whites. Thus, on this non-cognitive, motivation-sensitive test there is no evidence that Negro children perform less well than white children; if anything, just the opposite is true. The same tester, in the same session, also administered a standard verbal and nonverbal intelligence test to all these children. The average white-Negro differences in sigma units (based on the white <u>SD</u>) were 1.63 for the verbal IQ and 1.70 for the nonverbal IQ (Jensen & Rohwer, 1970, pp. 55-71). (These differences are equivalent to about 26 and 27 IQ points.)

It is sometimes claimed that lower performance on IQ tests results from poor attention, distractability, carelessness, inability to follow directions, and the like. So a test was devised to measure these factors independently of intellectual ability <u>per</u> <u>se</u>. The test makes no demands on knowledge, <u>g</u>, or memory. It is called a Listening-Attention Test. It is administered in the classroom by means of a tape recorder. High scores on the Listening-Attention test indicate that the subject is able to hear and distinguish correctly the numbers spoken by the voice on the tape, and to follow directions, keep pace with the examiner, and mark the answer sheet properly. The procedure is quite simple. The child is provided with a two-page answer booklet containing columns of paired numerals, 10 pairs to a column; each column is headed by a capital letter, alphabetically beginning with A. A clear male voice from the recorder says, "Put the point

of your pencil on the letter A. Now, I am going to say one number in each pair, and you should cross out the number I say -- cross it out with an X. Ready? 2 - 4 - 8 - 9 - 3 - ," etc. The numbers are spoken at a 2-second rate. At the beginning of each series (10 in all), the subject is told to put his pencil on the letter at the top of the list. This test, too, was given to all 4th, 5th, and 6th graders (1423 whites, 1214 Negroes) in an urban school system. There was no significant white-Negro difference in mean scores on this test at any grade level (Jensen & Rohwer, 1970, pp. 58-60).

It might be argued, however, that children perceive whether a test _is_ or _is not_ really an intelligence test, no matter what the examiner says or how it is labeled, and that this recognition depresses the performance of the Negro pupils. This, too, can be experimentally controlled.

A technique that lends itself ideally to this purpose is the free recall of uncategorized and categorized lists, abbreviated FRU and FRC, respectively. The FRU procedure consists of showing the subject twenty familiar and unrelated objects (_e.g._, ball, book, brush, toy car, gun) one at a time, and after the whole set has been thus exposed, asking the subject to recall as many of the items as he can remember. The same procedure is repeated for five trials, each time presenting the items in a different random order. The subject's score is the total number of items he recalls correctly on each trial; the items may be recalled in any order that they come to mind. This kind of rote memory, it has been found, shows little or no correlation with IQ. But by a seemingly little change in our set of items, we can turn this procedure into an intelligence test showing a very substantial correlation with standard IQ tests (Glasman, 1968). This is the FRC procedure, which is exactly the same as FRU as regards instructions and requirements of the task. But in FRC the lists are composed of items which

can be grouped into several conceptual categories, such as furniture, vehicles, clothing, tableware, etc. The single items, however, are always presented in a random order on each trial without reference to their conceptual categories. The same subjects are never given both FRU and FRC, so there is no basis for any subject's perceiving one test as being different from the other. Subjects are assigned at random to either the FRU or the FRC test. Both groups have the same examiner, the same instructions, and to all outward appearances the two tests do not differ in content, difficulty, purpose, or demands made upon the subject. There is no reason whatsoever that FRU and FRC should elicit different test-taking attitudes or motivational states. However, subjects who do not spontaneously tend to "cluster" the items of the FRC list into conceptual categories in recalling them, perform no better on FRC than on FRU. The degree to which a subject "clusters" the items conceptually (a tendency which generally increases from the first to the fifth recall trial) is related to the amount he is able to recall. It is both this amount recalled and especially the conceptual clustering tendency itself which are correlated with IQ. When there is little or no clustering, there is also no appreciable correlation with IQ. It then becomes a test of sheer rote memory, which is psychologically quite different from the *g* factor of intelligence tests.

When the FRU and FRC procedures were given to groups of Negro and white 4th graders, what was found? First, there was a slight but nonsignificant ($p < .162$) difference between Negro and white scores (*i.e.*, total recall over five trials) in the FRU test (Jensen & Rohwer, 1970, pp. 103-118). On the FRC test, however, the recall score of the white children very significantly ($p < .014$) exceeded the Negro mean, by about one standard deviation, as shown in Figure 14.

Note that the Negro group's performance in the non-categorized lists are hardly distinguishable. Whites, however, had much better recall on the categorized as compared with the non-categorized lists. The greatest differences, however, were found in the clustering score for the FRC test. This score indicates the degree to which the subject conceptually clusters the items in their order of output in recall. The more items of the same category that are recalled adjacently in sequence, the higher is the clustering score. The clustering measure itself is made to be independent of the amount recalled. The Negro-white difference in overall clustering score was great ($p < .005$), and while the white group showed a marked regular increase in clustering from trial 1 to 5, there was no increase across trials in the Negro group.

It is impossible to account for the lack of a significant Negro-white difference in FRU and the marked difference in FRC in terms of differences in test-taking attitudes, motivation, and the like. The racial difference in this case is clearly attributable to the different cognitive processes involved in these tests.

There are other tests which do not look anything like most intelligence tests but which in fact correlate with IQ and also show significant Negro-white differences. The Harris-Goodenough "Draw-a-Man" test is an example. The child is merely led to believe this is a test of his drawing ability and is told to draw a man on a blank sheet of paper. The drawings are scored against age norms for their degree of maturity. We gave this test to classes of Negro and white pupils from kindergarten through 6th grade,

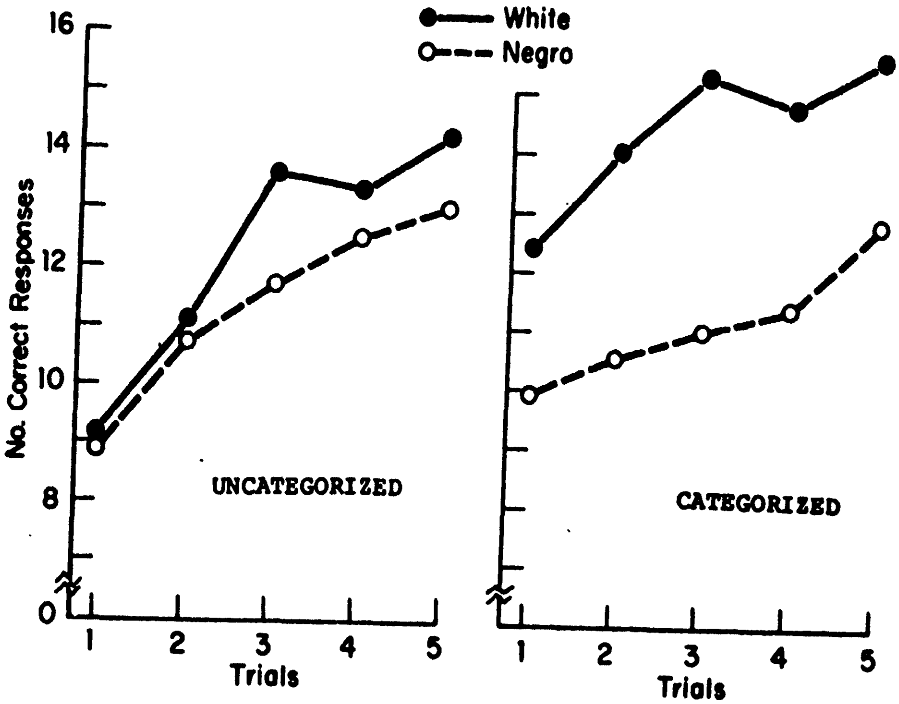


Figure 14. Amount of free recall of 20-item uncategorized and categorized lists.

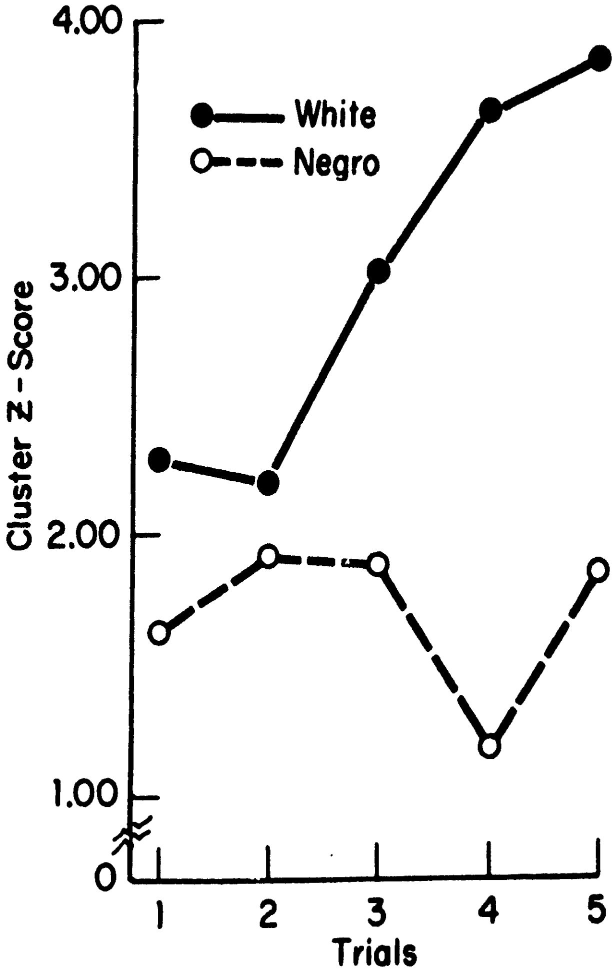


Figure 15. Amount of clustering in free recall of categorized lists.

removed all identification from the tests, and sent them away for scoring by an expert who knew nothing of the racial composition of the samples. At every grade there were substantial differences (the median difference being 12 IQ points) favoring the white pupils (Jensen & Rohwer, 1970, pp. 86-89). This test, however, is not a very good measure of **g** (as indicated by correlations of only .4 to .5 with Raven's matrices) and may be more culturally loaded than some tests of **g** which show larger Negro-white differences. (A review of other studies of Negro-white differences on the Draw-a-Man test is provided by Shuey, 1966, pp. 24-27; 83-84).

Language Deprivation

The use of dialect and often ungrammatical English in the speech of children called disadvantaged lends plausibility to the popular belief that these children's generally lower IQs and scholastic progress are attributable to environmentally caused verbal and lingustic deficits. We read that "Children from low socioeconomic groups develop deficits in intellectual functioning because they lack adequate intellectual, particularly verbal, stimulation . . . children in these groups receive less verbal stimulation from parents -- through being talked to, read to, taken on trips, etc. -- than children in middle-class groups, and that the parents are usually not very good examples for children to follow in learning language" (Furfey & Harte, 1970, p. 313). "It is to be expected that children from homes where certain words are used will do better on a vocabulary test involving those words than will children from homes where the words are never heard . . . most intelligence tests are loaded with middle-class content that is found to be more familiar to white children than to Negro children" (Brown, 1965, p. 186). Such statements do indeed appear very plausible, even self-evident.

But is linguistic deprivation actually an adequate explanation of intelligence differences? The point is not at issue that learning good

English is an advantage to upward social mobility. We are not concerned here with these secondary social consequences of grammar and dialect, but rather with the effect of language on intelligence and intelligence test scores. Several lines of evidence are highly relevant in evaluating the linguistic deprivation hypothesis of intellectual deficit.

In the first place, one would expect that if language differences played the predominant role in the lower intelligence test performance of Negroes, they should obtain their poorest scores on verbal tests and do relatively better on nonverbal and performance tests. In fact, just the opposite is most commonly found. The Wechsler-Bellevue and the Wechsler Adult Intelligence Scale, which are among the best individual tests of intelligence, consist of eleven subtests, six verbal tests and five performance tests, which yield a Verbal IQ and a nonverbal or Performance IQ. Every study of Negroes tested with the Wechsler scales reported in the literature, except for those involving nonrepresentative samples such as delinquents and prisoners, show higher Verbal IQ than Performance IQ (Shuey, 1966, pp. 295, 359-360, 371). On the Differential Aptitude Tests, Negro children in New York, whether they are middle-class or lower-class, were found to score higher on the Verbal ability test than on any of the other tests (Numerical, Reasoning, Spatial) (Lesser, Fifer, & Clark, 1965). The nationwide Coleman (1966) survey used verbal and nonverbal ability tests from grades 1 to 12 and found overall that Negro children did better (0.2σ to 0.3σ) on the verbal than on the nonverbal tests. (All other minority groups -- Puerto Rican, Indian, Mexican, and Oriental -- showed the opposite.[78]) Moreover, the verbal deprivation hypothesis of Negro IQ deficit should predict that the most disadvantaged Negroes with the lowest IQ -- those in the rural South -- should show a greater verbal deficit relative to their nonverbal test score than would be found in the comparatively more advantaged Negroes with higher IQs

in the urban North. But Coleman actually found just the opposite. The largest disparity between verbal and nonverbal scores, in favor of the verbal, showed up in Negroes of the nonmetropolitan South. Urban Negroes of the Northeast, Midwest, and Western regions, in fact, average two or three points <u>higher</u> on the nonverbal than the verbal tests beyond grade 3. Here, then, is a massive set of data which goes directly counter to the predictions of the verbal deprivation hypothesis: The presumably most deprived Southern Negroes actually do better on the verbal tests, the comparatively least deprived Northern Negores do better on the nonverbal tests. (On both verbal and nonverbal tests, the Northern and Urban Negroes excel over the Southern Negroes, but the disparity is <u>less</u> on the verbal tests. This appears paradoxical in terms of verbal-environmental deprivation theories of Negro intelligence.)

Do lower-class Negro children fail to understand white or Negro middle-class examiners and teachers, and even their own middle-class schoolmates, because of differences in accent, dialect, and other aspects of language usage? This proposition was examined in an ingenious experiment by Krauss and Rotter (1968). The groups they compared were low SES Negro children in Harlem and middle SES white children in the borough of Queens. Two age levels were used: 7-year olds and 12-year olds. Half the children in each group acted as <u>speakers</u> and half as <u>listeners</u>. The speaker's task was to describe a novel figure presented to him. The listener's task was to pick out this figure from a multiple-choice set of other figures solely on the basis of the speaker's description. The novel figures, drawn on cards, were nonrepresentational and were intentionally made difficult to name, so that they would elicit a wide variety of verbal descriptions. The speakers and listeners were paired so as to have every possible combination of age and race (or SES). (It must be rememberd that race and SES are

completely confounded in this experiment.) The score obtained by each pair of subjects was the number of figures the listener could correctly identify from the speaker's description. The results: the largest contribution to total variance of scores was the race (or SES) of the listener; the second largest contribution was the age of the listener. In other words, the 7-year old white (middle SES) children did better as listeners than the 1?-year old Negro (low SES) children. The speaker's age was the third largest source of variance. The race of the speaker, although a significant source of variance, was less than one-tenth as great as the race of the listener. In both age groups, the rank order of the mean scores for each of the four possible speaker-listener combinations were, from highest to lowest: white speaker-white listener; Negro speaker-white listener; white speaker-Negro listener; Negro speaker-Negro listener. The authors conclude: "...no support was obtained for the hypothesis that intra-status communication is more effective than inter-status communication" (Krauss & Rotter, 1968, p. 173). While these results seem paradoxical in terms of the linguist difference theories, they could be predicted completely on the basis of mental age obtained on a nonverbal intelligence test, such as Raven's matrices. The rank order of the mean of all possible race x age combinations of speakers and listeners could be predicted by the simple formula $MA_s + 2MA_L$, where MA is mental age, S is speaker, and L is listener. This is consistent with the hypothesis that it is intelligence rather than language usage *per se* which is the more important factor in communication. The results of several other studies of Negro-white differences based on speaker-listener interactions are consistent with this hypothesis and contradict the verbal deficit hypothesis (Harms, 1961; Eisenberg, Berlin, Dill & Sheldon, 1968; Weener, 1969; Peisach, 1965).

Does the disparity between a white middle-class examiner's standard English and the Negro child's ghetto dialect work to the disadvantage of the Negro child in a verbally administered individual IQ test such as the Stanford Binet? Quay (in press) attempted to answer this question by having a linguist whose speciality is the Negro dialect translate the Stanford-Binet into the Negro dialect. This form of the test was administered by two Negro male examiners to 50 4-year-old Negro children in a Headstart program in Philadelphia. Another 50 children, selected at random from the same Headstart classes, were given the test in standard English. The result: no significant difference (Negro dialect form was 0.78 IQ points higher than standard form). The author notes ". . . it is interesting that verbal items were passed with greater frequency than performance items. . ." "The analysis of item difficulty raises questions about the existence of either a language 'deficit' or a language 'difference' for Negro children having the experiences of the present Ss. At least their comprehension of the standard English of the Binet was not impaired."

Linguists and child psychologists who study the development of language are now finding, contrary to the popular belief, that Negro children, especially lower-class Negro children, are actually somewhat precocious in the most fundamental aspects of language development as compared with middle and upper-middle class white children. Baratz (1970), a sociolinguist and student of language development in Negroes, has argued that the Negro child is linguistically advanced compared to the middle-class white child. Entwisle (1970) has carried out a series of studies of the most basic aspects of language development based largely on children's free word associations. Children are asked to give "the first word you think of" in response to a standard set of stimulus words. The words represent different degrees of rarity and different grammatical form classes -- nouns, verbs, adjectives,

etc. The nature of the child's associations to the stimulus words follows a lawful developmental sequence. It is the type of response rather than the specific content of the response that is most important. Younger children, for example, are much more likely to give a syntactic phrase completion response to a stimulus word while older children will respond with another word of the same grammatical form class, called a paradigmatic association. For example, to the stimulus word cat the less mature child may respond with "drinks milk" or "nice"; the linguistically more mature child is likely to respond with "dog" (a noun of the same form) or "animal" (a noun of a supraordinate category). Or take the stimulus word begin. The less mature respond with "building a house," or "to cry" or "to eat," or "with" -- all responses showing a knowledge of how the word is used syntactically. Older children respond to begin with "start," "end," "stop," and the like. Children ages 4 and 5 are also more apt to give so-called klang associations than older children, e.g., begin -- "chin" or "lyn;" cat -- "sat" or "hat." Although preschoolers, of course, have never been exposed to formal grammar, their word associations and their speech reveal that they have already learned considerable grammar, as indicated by knowledge of pluralization, verb inflections, etc.

Now, what Entwisle (1970) has found in several studies is that in terms of these very basic developmental milestones in children's acquisition of language, low SES Negro children show a more precocious rate of development than middle SES white children. She writes: "Word associations of black and of white elementary-school children reveal, contrary to expectations, that slum children are apparently more advanced linguistically than suburban children at first grade..." But here is the interesting point: "White [italics added] first-grade slum children of average IQ give paradigmatic

responses to about the same extent as gifted (IQ 130) suburban children, and although inner-city blac! first-graders of average IQ lag behind inner city white first-graders they give more paradigmatic (i.e., mature) responses than white suburban first-graders of average IQ. Thus, at first grade the white child is slightly ahead of the black child when both are reared in the inner city, but the black slum child exceeds the white suburban child. The superiority is short-lived, however, for by third grade, suburban children -- whether blue collar or upper-middle-class -- have surpassed the inner city children, whether black or white. . . the temporary advance in linguistic development, and the subsequent decline, appears to be typical of the child in a poverty environment" (Entwisle, 1970). It may seem surprising that the rate of language development should so markedly decelerate in slum children between the first and third grades in school, where reading and cognitive enrichment are presumably being fostered. The nature of the changes from linguistic precocity to linguistic retardation in these children is interesting. Entwisle observes: "The relative developmental position of blacks and whites does shift with advancing age, however, and both inner-city blacks and whites show a slowed pace of development compared to suburban children by third grade. Again, however, the rate alone tells only a small part of the story, for while the semantic systems of white inner city children overlap considerably the semantic systems of white suburban children, semantic systems of black children depart significantly from both white groups, especially for more complex words." The main difference, pointed out in several examples mentioned by Entwisle, is that by fifth grade Negro children's responses in the word association procedure are more restricted to a very specific context, as contrasted with responses reflecting broader meaning, greater generality, conceptual categorizing, supraordination, and the like.

To explain these findings, Entwisle mentions such factors as the greater restriction of television viewing imposed on suburban preschoolers, lack of sufficient reinforcement for learning after school entry afforded to slum children, and a "lack of environmental forces to encourage semantic development." However, there is another possible interpretation of these findings which brings them theoretically under the purview of a much broader range of findings in developmental psychology. First, there is a fundamental biological principle, so general that it holds both across species and within a given species, which states that the more prolonged the infancy, the greater in general is the cognitive ability of the species at maturity. Precocity of early motor development, as assessed by infant tests, is negatively correlated with IQ at maturity among whites; and low SES white children, who show higher than average motor development scores in the first year of life, obtain below-average IQs as teenagers (Bayley, 1966). At birth and during the first year and a half of life, Negro infants, whether born in Africa or America, are physically and motorically more advanced than white infants; the majority of studies have shown this (Ainsworth, 1963, 1967; Bayley, 1965; Curti, 1935; Durham Education Improvement Program, 1966-67 a,b; Faladé, 1955, 1960; Géber, 1956, 1958a,b, 1960, 1962; Géber & Dean, 1957a,b, 1958, 1964, 1966; Gilliland, 1951; Kilbride, 1969; Knoblock & Pasamanik, 1953, 1958; Liddicoat, 1969; Liddicoat & Koza, 1963; Massé, 1969; Moreigne, 1962; Naylor & Myrianthopoulos, 1967; Nelson & Dean, 1959; Pasamanik, 1949; Ramarosaona, 1959; Scott, Ferguson, Jenkins, & Cutler, 1955; Vouilloux, 1959; Walters, 1967; Williams & Scott, 1953). Only three studies, in African samples, have reported findings of no significant differences in infant development (Falmagne, 1959; Langton, 1934-35; Theunissen, 1958).[79] Finally, language development is a species specific characteristic, peculiar but universal

to humans, which is intimately related to other developmental processes, including motoric behavior. Precocity of language development, as contrasted with the later role of language as a vehicle for abstract, conceptual processes, thus may be viewed as another reflection of generally accelerated sensorimotor development. The generally negative correlation between rapidity of early development and later level of cognitive ability reflected in intelligence tests and other indexes of conceptual ability is consistent with the later deceleration of linguistically precocious children -- a deceleration that shows up at the age when the child's language begins to reflect more complex, abstract, and conceptual mental processes.

Students of language development recognize that it is largely under the control of innate factors. Houston (1970), a psycholinguist, points out that ". . . all children learn language merely by being placed in the environment of the language and. . . they do not need any special training or conditioning whatever to achieve this (4 references). Further, all children appear to learn language in about the same length of time, namely, from four to six years . . . Given the open-ended variation in learning environments previously noted and given the lack of directed reinforcement for language or other behavior in children characteristic of many societies, the argument for a biological basis for language acquisition is convincing . . . It is now believed by linguists that man has an innate biological capacity for language acquisition, a capacity which has been described as a species-specific and species-uniform language-acquisition device which functions uniquely in the language-acquisition process and the operation of which is constant for all children. Various biological and neurophysiological correlates of the language-learning process have been discovered, so that this position is strengthened" (pp. 949-950). In these aspects, language acquisition can be likened to the child's learning

to walk. All physically normal children learn to do so, given the barest opportunity, and they go through the same sequence from crawling, creeping, and toddling to walking, although showing slight individual differences in rates of acquisition at each stage along the way. Houston points out that ". . . language-acquisition stages seem invariant; it should be additionally noted that all children have rules by which they produce their language at each stage of the acquisition process, irrespective of the particular language or form of language they are acquiring (3 references)" (p. 951).

Is the later lag in cognitive development seen in low SES children and especially in low SES Negro children, between grades 1 and 3, due to delayed effects of verbal deprivation during the preschool years or to insufficient verbal stimulation outside of school? Does language deficiency *per se* hinder conceptual and abstract thinking? In seeking answers to these questions, it should be instructive to study the most verbally deprived children we know of -- children who are born totally deaf. Since the year 1900 there have been some 50 comparative studies of the intelligence of the congenitally deaf; all these studies have been reviewed and summarized in two articles by McCay Vernon (1967, 1968). As might be expected if deafness constitutes a severe form of verbal and language deprivation, congenitally deaf children score well below normally hearing children on strictly verbal tests. At age of school entry, when normal children have a vocabulary of 2,000 to 8,000 words and a well-developed syntax, deaf children usually know absolutely no words at all, and it is only after about four years of education, at about ten years of age, that these children can begin to compete with the average first grader in vocabulary and other language skills; about 35 percent of such children never achieve functional literacy, so great is their verbal handicap.

But how do these deaf children score on nonverbal performance tests of intelligence? Vernon summarizes his review of all the literature on this

point: ". . .the research of the last fifty years which compares the IQ of the deaf with the hearing and of subgroups of deaf children indicates that when there are no complicating multiple handicaps, the deaf and hard-of-hearing function at approximately the same IQ level on performance tests as do the hearing" (1968, p. 9.)[80] Note that the pattern of test scores for the deaf is just the opposite to that of Negro children, who do better on the verbal and poorer on the performance tests. Vernon concludes that there is no functional relationship between verbal language and cognition; verbal language is not the mediating symbol of thought, although verbal behavior may serve to mediate and express thought processes in ways I have explicated in detail elsewhere (Jensen, 1970). Another student of the psychology of the deaf, Hans Furth (1964), has come to similar conclusions: "In summary, then, the reported investigations [of the cognitive abilities of the deaf] seem to emphasize as legitimate the distinction between intellective and verbal skills. The ability for intellective behavior is seen as largely independent of language and mainly subject to the general experience of living. Various sources of empirical evidence confirm the theoretical position that just as language learning is not closely related to intellectual endowment so intellective performance is not directly dependent on language" (p. 162). Specifically, Furth concludes, "Empirical studies of deaf people's performance on nonverbal cognitive tasks were reviewed. Deaf were found to perform similarly to hearing persons on tasks where verbal knowledge could have been assumed a priori to benefit the hearing. Such evidence appears to weaken a theoretical position which attributes to language a direct, general, or decisive influence on intellectual development" (p. 145).

Another important difference between low SES children and children who are verbally deprived because of deafness is that while the former begin to

lag in linguistic and intellectual development after beginning school, the latter show a gradual catching up to the average level as they progress in school -- it merely takes them longer to acquire information because of their severe sensory handicap. But once it is acquired, normal mental development ensues. A study of the developing conceptual capacities of the deaf concluded " . . . the differences found between deaf and hearing adolescents were amenable to the effects of age and education and were no longer found between deaf and hearing adults. Dissociation between words and referents, verbalization adequacy, and (conceptual) level of verbalization were not different for deaf and hearing subjects. Our experiments, then, have shown few differences between deaf and hearing subjects. Those found were shown to fall along a normal developmental line and were amenable to the effects of increased age and experience, and education" (Kates, Kates, & Michael, 1962, pp. 31-32).

Thus the language deprivation theory of the Negro-white IQ difference simply does not accord with the facts, and so we must turn to other possible explanations.

Culture-Biased Tests

The claim that intelligence tests are culturally biased in favor of white middle-class children and are therefore invalid when applied to minority children (or to lower-class white children) is undoubtedly the commonest argument against studies of subpopulation differences. The SPSSI Council (1969) states: "We must also recognize the limitations of present-day intelligence tests. Largely developed and standardized on white middle-class children, these tests tend to be biased against black children to an unknown degree. While IQ tests do predict school achievement, we cannot demonstrate that they are accurate as measures of innate endowment. Any

generalizations about the ability of black or white children are very much limited by the nature of existing IQ tests." This view has been the basis for moves to abolish all testing in the public schools. Thus we read in a newspaper: "The Board of Education voted unanimously last night to appoint a special committee to decide whether all psychological testing of minority children should be stopped. The resolution cameafter a score of black community leaders pleaded for an immediate moratorium on achievement and intelligence testing of minority children. They said the tests, designed for middle class whites, were invalid for minority groups." (San Francisco Chronicle, May 6, 1970, p. 18). The news report quoted a psychologist as saying, "Asking a black child the advantages of having a checking account when most black families don't have them is about as fair as asking white children about chitterlings when most white families don't eat them." The public is left with the clear impression that all intelligence tests are comprised of questions of factual information typical of what children in middle or upper class homes are most likely to learn and children from poor homes are least likely to learn. One can always point to some items or some tests which seem to illustrate this point. The next step is to brand intelligence tests as instruments of social injustice, devised and used by the Establishment to maintain the social class structure of the society. Thus an educational sociologist writes: "In view of the close relationship between IQ scores and social class in Big City, it seems that one very destructive function of the IQ score is that it serves as a kind of cement which fixes students into the social classes of their birth. IQ is the supreme and unchallengeable justification for the social class system" (Sexton, 1961, p. 51). This overlooks the fact that more than a third of the population changes social status each generation and that the correlation between SES

and IQ is much higher for parents than for their children. Actually, IQ tests, much more so than interviews, teacher's impressions, and school grades, can have a liberalizing influence on the education and upward mobility of lower-class children, since good IQ tests can "read" through the superficial veneer of cultural advantages related to social status. Many intellectually gifted children who might otherwise go undiscovered by their parents, peers and teachers are found by means of intelligence tests. As sociologist Otis Dudley Duncan (1968, p. 11) put it: ". . . intelligence contributes a large share of variance in achievement (i.e., education, occupation, income) that is unrelated to the social class of birth. . . in view ofthe loose relationship between IQ and social class in the United States, it seems that one very constructive function of the ability measured by intelligence tests is that it serves as a kind of springboard, launching many men into achievements removing them considerable distances from the social class of their birth. IQ, in an achievement-oriented society, is the primary leaven preventing the classes from hardening to castes." It should also be pointed out that individual differences in the ability measured by IQ tests exist and contribute to educational and occupational achievement and to social mobility, whether or not measured by IQ tests. The doctor will not alter his patient's fever by throwing away the thermometer.

The last refuge of the critics of IQ tests is to argue that there is no such thing as intelligence and that even if it did exist, it could not be measured. A more sophisticated version of this argument is that there are so many different kinds of abilities that it is meaningless to speak of intelligence as a general ability or a g factor which is relevant to a wide variety of achievements. It is an exercise in futility to attempt to support these arguments in the face of all of the evidence which clearly refutes them. Readers who wish to see these issues trenchantly spelled out

are referred to Quinn McNemar's (1964) presidentail address to the American Psychological Association. Philip E. Vernon (1965, p. 724) has put it most succinctly: "A general intelligence factor seems unavoidable since substantial positive inercorrelations are found when any cognitive tests are applied to a fairly representative population." One of the many practical consequences of this fact is noted, as Vernon continues: "When I visited some military psychological establishments in 1957, I was told more than once that military psychologists could not ignore g. Try as they would to find differential tests for different army trades, intercorrelations were so high that recruits appeared to be differentiated more by all-round level of ability than by type of ability, that is to say, by g rather than by factor profile."

Standardization and Predictive Validity

First, let us look at the standardization argument, typically expressed as it was in the New Scientist (July 23, 1970): "Is it surprising that Negro children do badly on culturally loaded tests standardized only on white children?"

All that standardization means is that the test has been given to some fairly large and representative sample of some population, and the distribution of raw scores (i.e., the number of test items the subject gets "correct") is converted into standard scores,[81] so that each age group in the population sample will have the same mean and standard deviation. (On most IQ tests conventionally the mean is set at 100 and the SD at 15.) If the raw scores do not conform to the normal or Gaussian bell-shaped distribution, they may be normalized by converting them to percentile scores which are then expressed as deviates of the normal distribution. In essence, standardization is the process of re-scaling raw scores in terms of the mean and standard deviation

of the so-called normative population. The normative population may or may not include one or another racial subpopulation. And some normative samples are more representative of one geographical region or its racial composition than another. It is possible to obtain separate norms for whites and Negroes, and it is possible to have norms based on the combined groups each represented proportionally to their frequency in the general population. Various standardized tests have used one or another of these methods. The important point, however, is that it makes absolutely no real difference in terms of the rank order of individuals or of subpopulations on the test, and it has no effect whatsoever on the predictive validity of the test, any more than it lowers a patient's temperature to change from a Fahrenheit to a Centigrade thermometer. We are simply assigning different numbers to the same relative differences. For example, a test standardized exclusively in the white population and given a mean of 100 and a SD of 15, will show a mean of approximately 85 and a SD of 13 when given to a representative sample of the Negro population. If we standardize the test in a combined sample of whites and Negroes, represented in the proportions of their frequencies in the general population of the United States (approximately 89% whites, 11% Negroes), the general mean IQ will still be set at 100 and the SD at 15. But on this scale the Negro mean will be 87.1 and the white mean will be 101.6, for a difference of 14.5 IQ points. But the percentage of Negroes exceeding the white median (i.e., median overlap) will remain exactly the same, viz., 14.2%. (Median overlap has this advantage as a measure of group differences — it is invariant regardless of the scale of measurement.) Thus, it is apparent that re-standardizing or re-scaling IQ tests would make no essential difference in Negro-white comparisons.

Having separate sets of norms for Negro and white populations, each given a mean of 100 and SD of 15, could only impair the predictive validity

of the tests when they are used in mixed populations. This is because intelligence tests standardized on whites or on Negroes and whites together have the same predictive validity for Negroes as for whites (see Jensen, 1971d and Stanley, 1971, for reviews of this evidence). In other words, the tests themselves are color blind and predict the same scholastic and occupational performance for individuals obtaining the same score, regardless of race. (When prediction discrepancies have been found, they have usually been "in favor" of Negroes, i.e., the Negroes perform less well on the criterion than was predicted by the test.) Jensen (1971) found in a large elementary school sample ($\underline{N}$ = 6,569) that including "race" (white, Negro, or Mexican) along with IQ and other psychological tests in a multiple regression equation for predicting scholastic achievement adds no significant increment to the prediction (i.e., $\underline{R}^2$). Also, factor analyses of a large battery of mental ability tests yield the same factors in Negro and white samples, although the specific factor loadings on some tests may differ for Negroes and whites. In summary, ability tests behave very much the same in Negro and white populations. This would not be the case if IQs were based on separate norms for the two populations. As it is, a Negro child and a white child with the same IQ can be expected to perform equally well in school or on the job (in so far as it depends upon intellectual ability). If there were different norms for the two groups, a Negro person and a white person with nominally the same score would not be expected to perform equally in school, etc. In short, in terms of their predictive validity, that is to say their practical usefulness, most standard IQ tests, scholastic aptitude tests, and the like, are not biased against Negroes.[82]

<u>The Culture-Bias Hypothesis</u>. Though "culture-bias" is the term most commonly used, a better term would be "status bias," since in making group comparisons within a given culture we are concerned with test content biases that may discriminate according to the differential experiences of persons who have grown up in different social strata. When Negro and white children

grow up in the same locality, attend the same schools, watch the same television programs, have the same toys, eat the same food bought in the same stores, wear the same clothes, etc., they have much more in common culturally than not, as judged on a world-wide scale of cultural diversity. Negro and white children growing up, say, in Berkeley, California, surely have much more culturally in common with one another than either group has in common with, say, the Eskimos or the Australian Bushmen. Whatever experiential differences exist are largely social status differences rather than cultural differences in any meaningful sense of the word.

To say that an intelligence test is culturally biased or status biased means that the knowledge, skills, and demands of the test sample the specific learning opportunities of one subpopulation (*i.e.*, social class or racial group) more than of another. One can think of many examples of questions that would be easier for children of one subpopulation than of another. The question, for example, "What are chopsticks made of?"would favor Oriental children; "What are tortillas?" would favor Mexican children; "What are chitterlings?" would favor Negro children, "What is the Talmud?" would favor Jewish children, and so on. One could presumably devise a test composed entirely of specially selected items that would give a marked advantage to any particular subpopulation one might choose. The culture-bias hypothesis claims that this in fact is what has been done: intentionally or unintentionally, standard intelligence tests have been composed of items which favor white middle- and upper-class whites and disfavor all other groups, especially Negroes, other minorities, and lower-class whites. How valid is this claim? Since I have discussed these issues at length elsewhere (Jensen, 1968c, 1970b), I will here attempt only a brief summary of the main points.

Surely one can point to "culture-loaded" items on many standard intelligence tests. Questions about exotic zoo animals, fairy tales, and musical

instruments are obvious examples of items that should favor children from well-to-do homes and disfavor children from poor homes which afford little opportunity to learn about such things. In one obvious attempt to discredit IQ tests, five highly "culture-loaded" items from the Comprehension subtest of the Wechsler Adult Intelligence Scale (WAIS) were selected for display by proponents of the culture-bias hypothesis of Negro-white IQ difference (Bulletin of the Cambridge Society for Social Responsibility in Science, July 18, 1970, p. 6). It may therefore seem ironic to discover that, in fact, among the 11 subtests of the WAIS, the one on which Negroes actually differ least from whites is the very Comprehension test that was held up as an example of test items that might seem to be culturally biased against Negroes (*e.g.*, "Why should people pay taxes?" "Why does land in the city cost more than land in the country?" "Why are laws necessary?") (Plotkin, 1971, p. 6; Shuey, 1966, p. 407). Negroes actually score lowest on the Block Design subtest, a nonverbal test requiring the subject to copy patterns of increasing complexity with a set of 16 colored one-inch blocks. This is probably the least culture-loaded subtest of the WAIS. This is not a result peculiar to the Wechsler tests. In general, Negroes obtain higher scores on tests which by any reasonable criteria appear to be more culture loaded than on items that are less culture-loaded. This has been demonstrated most dramatically in a study by McGurk (1951, 1953a, 1953b, 1967). He compared the performance of Negro and white 18-year-old high school students on highly culture-loaded as compared with minimally culture-loaded intelligence test items. For this purpose, to quote McGurk (1967, p. 374), "A special test was constructed, half the questions of which were rated as depending heavily on cultural background (the culture questions) while the other half were rated as depending little on cultural background (the non-cultural questions). Each set of questions yielded a score -- either a

culture score or a non-culture score. McGurk found that the "Negroes performed better (relative to the whites) on the culturally loaded questions" (1967, p. 378). This comparison was based on Negro and white groups selected in such a manner that "Negroes and whites were paired so that the members of each pair -- one Negro and one white -- were identical or equivalent for fourteen socio-economic factors" (McGurk, 1967, p. 379).

How can we understand such seemingly paradoxical results, which are the rule and not the exception? In order to find the answer, I have carried out item analyses of many kinds of intelligence tests, seeking those which discriminate the most and the least between Negro and white subjects, as well as between white lower and middle SES subjects. When one brings together large numbers of test items solely on the basis of whether they discriminate minimally or maximally between Negro and white (or low and middle SES) samples, the answer to the paradoxical findings becomes apparent. All intelligence tests are intentionally devised so that the items vary in difficulty, usually beginning with the easiest items and increasing gradually to the most difficult items. Item difficulty is objectively defined simply in terms of the percentage of the normative population that fails to give the correct answer to the item. Examination and statistical analyses of a wide variety of test items reveals that items are graded in difficulty along two main dimensions (not mutually exclusive). One dimension is rarity or infrequency of opportunity to learn the content of the item. Many general information items and vocabulary items vary in difficulty along this rarity dimension, _e.g._, "What is the Bible? _vs_. "What is the Koran", and define "physician" _vs_. "philologist." It happens that the type of items that increase in difficulty along the rarity dimension are those we call the most culture-loaded. Their difficulty depends upon their rarity rather than upon the _complexity_ of the mental processes required for arriving at the correct

answer. Complexity is the other main dimension along which test items increase in difficulty. Items differ in the amount of mental manipulation and transformation of the elements of the question that they require in order to arrive at the correct answer. Thus, the question "What is the color of fire engines?" is low on both rarity and complexity, while the question "If a fire engine can go no faster than 50 miles per hour, how long will it take it to get to a fire five miles away?" is also low in rarity but considerably higher in complexity. Similarly, digit span memory (repeating a series of numbers after hearing the series spoken once by the examiner) is low on complexity as compared with number series completion, e.g., "What number should come next in the series: 1, 2, 4, 5, 7, __?__." A Wechsler Information test item like "Who wrote the Republic" is difficult because of its rarity. This Wechsler-type Arithmetic item on the other hand, may have exactly the same difficulty level, but it is difficult because of its complexity: "Six men can finish a job in four days. How many men will be needed to finish it in half a day?" Negroes do much better, relative to whites, on the Information than on the Arithmetic subtest, despite the fact that the Information items are more culture-loaded. Though it surely is not a necessary condition, it happens that in most intelligence tests there is an inverse relation between items standing on the rarity dimension and on the complexity dimension. The rarest, most culture-loaded items involve the least complexity, and the most complex items involve the most common contents. And what we find is that the degree to which items discriminate between social classes and between Negroes and whites is much more a function of the item's complexity than of its rarity or culture-loading. This is true whether the complexity involves verbal, numerical, or spatial materials. The degree to which test items call for mental manipulation, transformation, conceptualization, and abstraction --

and not so much the rarity or culture-loading of their contents -- is what mostly determines the Negro-white discriminability of test items. On the other hand, some subpopulations -- American Orientals, for example -- show just the reverse; they do relatively better (usually exceeding the white population) on those items most heavily loaded on the complexity dimension. Orientals are somewhat disadvantaged on tests to the extent that cultural items are included as opposed to complexity items, while just the opposite typically is true for Negroes.

When many test items of various types are included in a factor analysis, the degree to which they are loaded on the *g* factor (*i.e.*, the ability factor which is common to all intelligence tests and mainly accounts for their intercorrelations) is related more to the complexity of the items than to the rarity of their contents, especially if the tests are given to a culturally and socio-economically heterogeneously sample.) That is to say, the items that increase in difficulty along the complexity dimension better represent the *g* factor of intelligence in a heterogeneous population than do the more culture-loaded items.

By minimizing rarity and maximizing the varieties and degrees of complexity, it is possible to produce tests which are relatively "culture-fair" or "culture-reduced." No one claims that there is any test which is perfectly "culture-free," and so to attack "culture free" tests is to attack a staw man, a favorite occupation of many enviornmentalists. "Culture free" is an idealized and unattainable end-point of an actual continuum along which various tests (or test items) can in fact be rank ordered. To say there is no such thing as a "culture free" test does not mean that tests cannot be ordered along a dimension (or a number of dimensions) representing the degree to which they utilize contents having differential rarity in

various subpopulations. Just as there are no "perfectly soft" or "perfectly hard" gems, it is nevertheless possible to rank gems along a soft-hard continuum.

Numerous attempts have been made to devise culture-reduced tests, the main approaches to which I reviewed elsewhere (Jensen, 1968c). All approaches have been essentially an attempt to minimize the rarity factor, either by using content that is equally common to all status groups within the culture or non-representational content that is equally unfamiliar to everyone. Item difficulty then is controlled by the complexity of the mental operations with the equally familiar materials needed to find the answer. One of the pioneering attempts at this, now of historical interest, is the defunct Davis-Eells Games, developed in 1951. The items, represented as games, were cartoons of children doing ordinary things in very familiar settings; in fact, the settings were more typical of a lower-class environment than of a middle-class environment. No reading was required and the tests were untimed -- features thought to favor lower SES relative to middle SES children. Practical judgment and common sense inferences are called for in solving most of the problems. One cartoon, for example, shows three panels, each depicting a boy trying to get over a high backyard fence. One boy is piling up boxes and rubbish cans in a most unstable fashion, one is futily jumping, and one is stacking boxes in a stable fashion. The testee simply checks the picture he thinks shows the best method for getting over the fence. But in order to increase item difficulty with such familiar materials, the problem situations had to be made increasingly complex in the inferences and judgments called for in order to solve them. Logical reasoning was needed, though it always involved only commonplace practical situations. But the test was entirely unsuccessful in the view of those who had hoped

it might eliminate social class and racial differences in mean performance. Group differences approximately equal to those found with the ordinary standard IQ tests were found with the Davis-Eells Games. Since they essentially failed in their main purpose and had certain psychometric defects as well, the Games were dropped. Subsequent attempts along the same lines were made by Davis and his co-workers, using items believed to be intrinsically motivating, similar to real-life situations, equally familiar to all social classes, and without time limits. But these tests, too, yielded lower scores for Negroes, lower, in fact, than found for low-status whites. As of the present time, no one yet has succeeded in devising a test that does not discriminate between representative samples of Negroes and whites and which also can be shown to have any *g* loading (which is essentially the complexity factor) or any validity in terms of correlation with any external educational or occupational criteria. If group differences were due to cultural bias in the test and not to true differences in intelligence, it should be possible to devise culturally appropriate tests that eliminate the group difference and yet retain the tests' validity. The fact that no one has yet been able to devise a test, either culture-fair or culture-loaded, on which Negroes perform as well as whites or other minorities, despite many serious attempts to do so, is a strong argument against the culture-bias hypothesis. In this connection, it is interesting to note the kinds of changes that have had to be made in the Wechsler tests to make them have the same reliability, validity, and intersubtest correlations in various foreign countries as are found in the normative population of the United States. The necessary changes are surprisingly minimal. None of the Performance tests has to be changed except one involving a picture of the American flag, and the changes required in the Verbal tests are usually no more than translation into the foreign language, with but few exceptions, such as changing two or three Information

subtest items involving names of American historical figures and geographical features of the United States commonly taught in our schools. Equivalent items are easy to find for other countries (e.g., "What is the population of Japan?" "How far is it from Tokyo to Osaka?" etc.). When so few and superficial changes can make it possible for persons in foreign lands and different cultures to perform on a par (or exceed) white Americans, it is a wonder why the most diligent efforts have failed to devise an intelligence test on which American Negroes can score on a par with the rest of the population.

The most successful culture-reduced tests have been those employing simple figural materials, requiring subjects to engage in reasoning, inference, generalization, and other basic mental processes in terms of relationships between geometric forms, patterns, etc. Such tests are Raven's Progressive Matrices, Cattell's Culture-Fair Tests of *g*, the Lorge-Thorndike Nonverbal Intelligence Test, the Street Gestalt Test, the Gottschaldt Embedded Figures Test, and others. MacArthur and Elley (1963) set up certain desirable criteria for culture-reduced tests and studied a host of such tests along with conventional IQ tests to determine which of the many tests came closest to meeting their criteria. Raven's Progressive Matrices and Cattell's Culture-Fair Tests of *g* proved to be the best in this study, which showed that "(1) Culture-reduced tests sample the general intellectual ability factor (*g*) as well as or better than conventional tests. (2) Most culture-reduced tests show negligible loadings on verbal and numerical factors. (3) Culture-reduced tests show significantly less relationship with socio-economic status than do conventional tests. (4) A conventional test (California Test of Mental Maturity) showed a significant increase in relationship with socioeconomic status over four years, whereas the Progressive Matrices showed no change. (5) Verbal items from the CTMM showed greater

variation in item discrimination between social classes than did items from the Progressive Matrices" (p. 118).

The Progressive Matrices test has been used in numerous truly cross-cultural studies. These studies show mean differences between various ethnic and cultural groups the directions of which are not at all in accord with the popular notion that groups are handicapped on IQ tests directly in relation to their degree of environmental and cultural dissimilarity from that of the white middle and upper-middle class population of the United States or Western Europe. It would be hard to find an environmentally and culturally more dissimilar group than the Eskimos living in the icy wastes far above the arctic circle. Yet representative samples of these Eskimos score at or above white Canadian norms on the Progressive Matrices (MacArthur, 1968). Berry (1966) found Eskimos samples scoring near his Scottish samples (one of the highest normative groups) on the Progressive Matrices and the Embedded Figures Tests. Vernon (1965a), too, has found that on the Matrices and similar tests, such as the Kohs block designs and Abstractions, Eskimos and Canadian Indians score much higher than Jamaican Negroes. Vernon seeks an environmental explanation of the marked disparity: "Now economic conditions are extremely poor in all three groups (Eskimos, Indians, Jamaicans), and there is similar family instability and insecurity. Thus it seems reasonable to attribute the better performance of Eskimo and Indian groups to the greater emphasis on resourcefulness in the upbringing of boys, perhaps combined with their strong masculine identification. True, the traditional hunting-trapping life is rapidly disappearing and the majority of parents are wage earning or on relief, but the children are still brought up permissively and encouraged to explore and hunt. Moreover, a subgroup

of the Eskimos who came from the most isolated Arctic communities scored better on all three of the tests just mentioned (Matrices, Embedded Figures, Abstractions) than did those who lived in closer contact with whites and had become more acculturated" (Vernon, 1965, p. 732).

In the United States Negroes generally average about 1 SD or more below whites on the Matrices. Figure 12 (page 000) shows the Matrices scores of large representative samples of California school children at several grade levels (Jensen, 1971a). The scores are on a T scale with the overall mean at 50 and a SD of 10 points. The results are particularly interesting in view of the fact that on the Home Index, a measure of environmental and cultural advantages, the Mexican group in this study is as far below the Negro group as the Negro is below the white. Figure 16 shows scores of the same three groups on a composite rote memory test which is difficult, requiring sustained attention, concentration, and motivation, but makes no demands on reasoning or abstract conceptual abilities.

One of the most status-fair tests, at least for children who are in school and have had experience with paper and pencil, is the Figure Copying Test (see Figure 1, page 000). The child is asked merely to copy the ten forms, each on a separate page, while they are in full view, without time limit. The children's drawings can be scored with a high degree of reliability for correspondence to the model and for maturity of the drawing. In factor analyses carried out separately in white, Negro, and Mexican samples, this test has a substantial g loading in all groups, comparable to that of Raven's Matrices. The test scores of kindergarten children also are prognostic of readiness for the traditional school learning tasks of

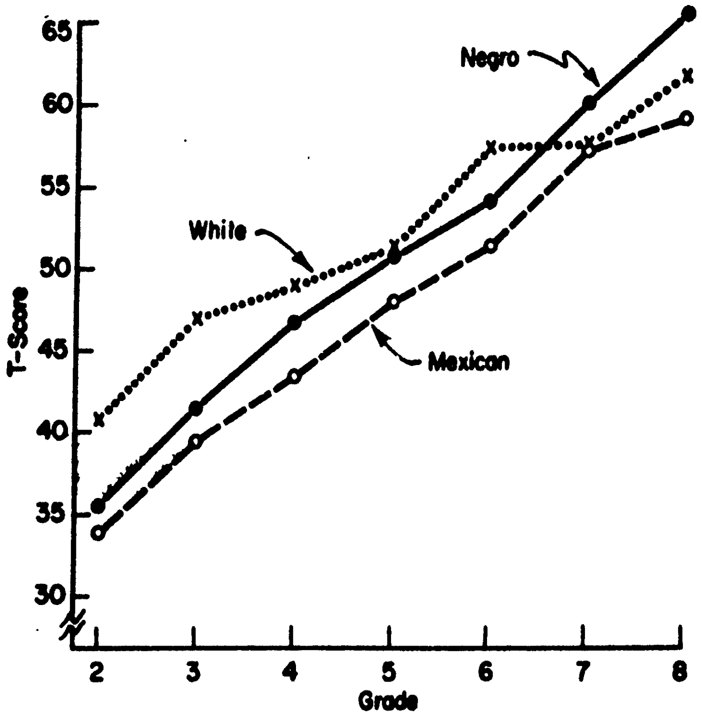


Figure 16. Mean T scores ($\bar{x}$ = 50, SD = 10) on composite rote memory tests. (From Jensen, 1971a)

the primary grades. The high level of motivation elicited by this test is indicated by the fact that the minimum score obtained in each group at each grade level increases systematically. This suggests that all children are making an attempt to perform in accordance with the instructions. Also, virtually 100 percent of the children in every ethnic group at every grade level attempted to copy every figure. The attempts, even when totally unsuccessful, show considerable effort, as indicated by the re-drawing of the figure, erasures and drawing over the figure repeatedly, in order to improve its likeness to the model. It is also noteworthy about this test that normal children are generally not successful in drawing figures beyond their mental age level and special instruction, coaching and practice in drawing these figures hardly improves the child's performance. Figure 17 shows the scores on this test of several ethnic and social class groups

totalling nearly ten thousand children in kindergarten to 4th grade in 21 California schools. The four ethnic groups are Oriental (O), White (W), Mexican (M), and Negro (N). The letter "U" represents schools in an urban, relatively upper status community socio-economically as compared with the average school district in California; "L" represents schools in comparatively lower status rural districts. The groups are ranked on a composite index of socioeconomic status (SES), with SES 1 as the highest, representing largely professional and business-managerial upper-middle-class families. Note that the rank order of SES does not strictly correspond to the rank order of performance in Figure Copying. The Orientals exceed all other groups, and the Mexicans, who are at the bottom in SES, score only slightly below the whites. At fourth grade the range of group mean differences on

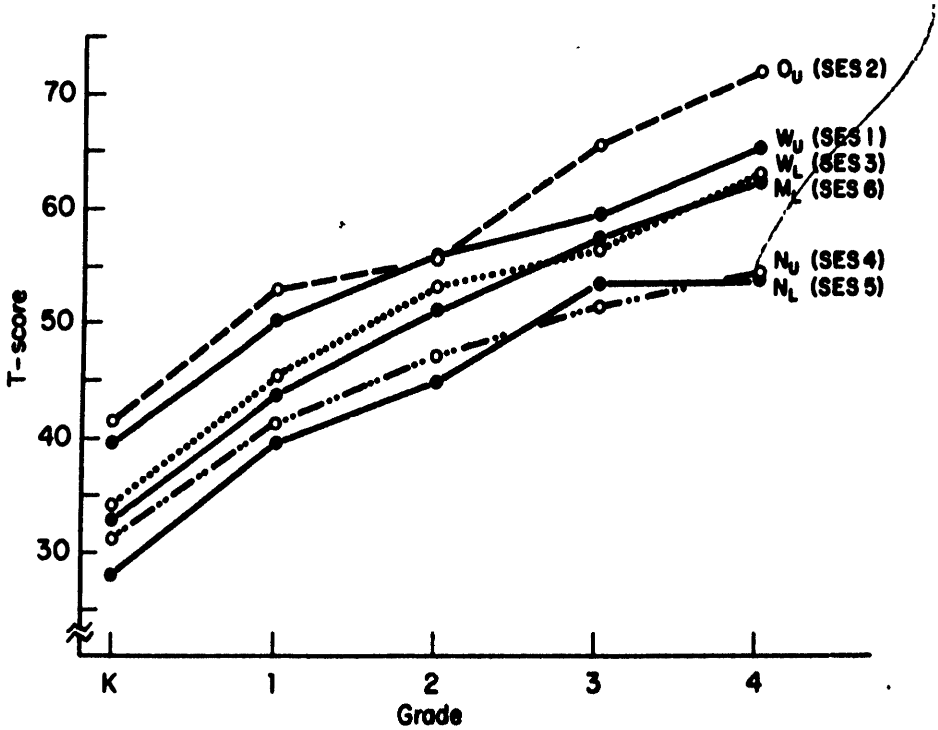


Figure 17. Oriental (O), white (W), Mexican (M), and Negro (N) groups from socioeconomically urban, largely middle to upper-middle class (U) and rural, largely lower to middle class (L) communities. The six groups are ranked from highest (SES 1) to lowest (SES 6) on a composite index of socioeconomic status.

the test spans more than 2 SDs. Negro fourth graders, on the average, match the performance of Oriental children in the first grade. These findings are consistent with results obtained at Yale's Gesell Institute using a battery of similar developmental tests with Negro and white elementary school children (Ames & Ilg, 1967). Especially for children who have been exposed to three or four years of schooling, such marked differences in performance would seem most difficult to explain in terms of differential experiences, motivation, and the like.

I have suggested previously that tests which are more culture-fair or status-fair can be thought of as having higher heritability in an environmentally heterogeneous population than highly culture- or status-loaded tests (Jensen, 1968c, pp. 81-86). Evidence from kinship correlations on various tests is consistent with this formulation.

Now, if we accept this premise that a test's culture-loading is inversely related to its heritability in a given population, let us examine the consequences of comparing the regression of a culture-loaded test upon a hypothetical culture-free test, and vice versa, in each of two hypothetical populations, A and B. If differences are found between groups A and B, one of three hypotheses can be invoked to explain the difference: (1) the groups are genetically equal but differ environmentally; (2) the groups are environmentally equal but differ genetically; or (3) the groups differ both genetically *and* environmentally. The consequences of each hypothesis are shown in Figure 18. Our hypothetical perfectly culture-free or environment-free (meaning $h^2 = 1$) test measures the genotype, G; the culture-loaded test measures the phenotype, P. (The phenotypic value, P, is the sum of the genetic and enivronmental values, *i.e.*, $P = G + E$.) Assume that the heri-

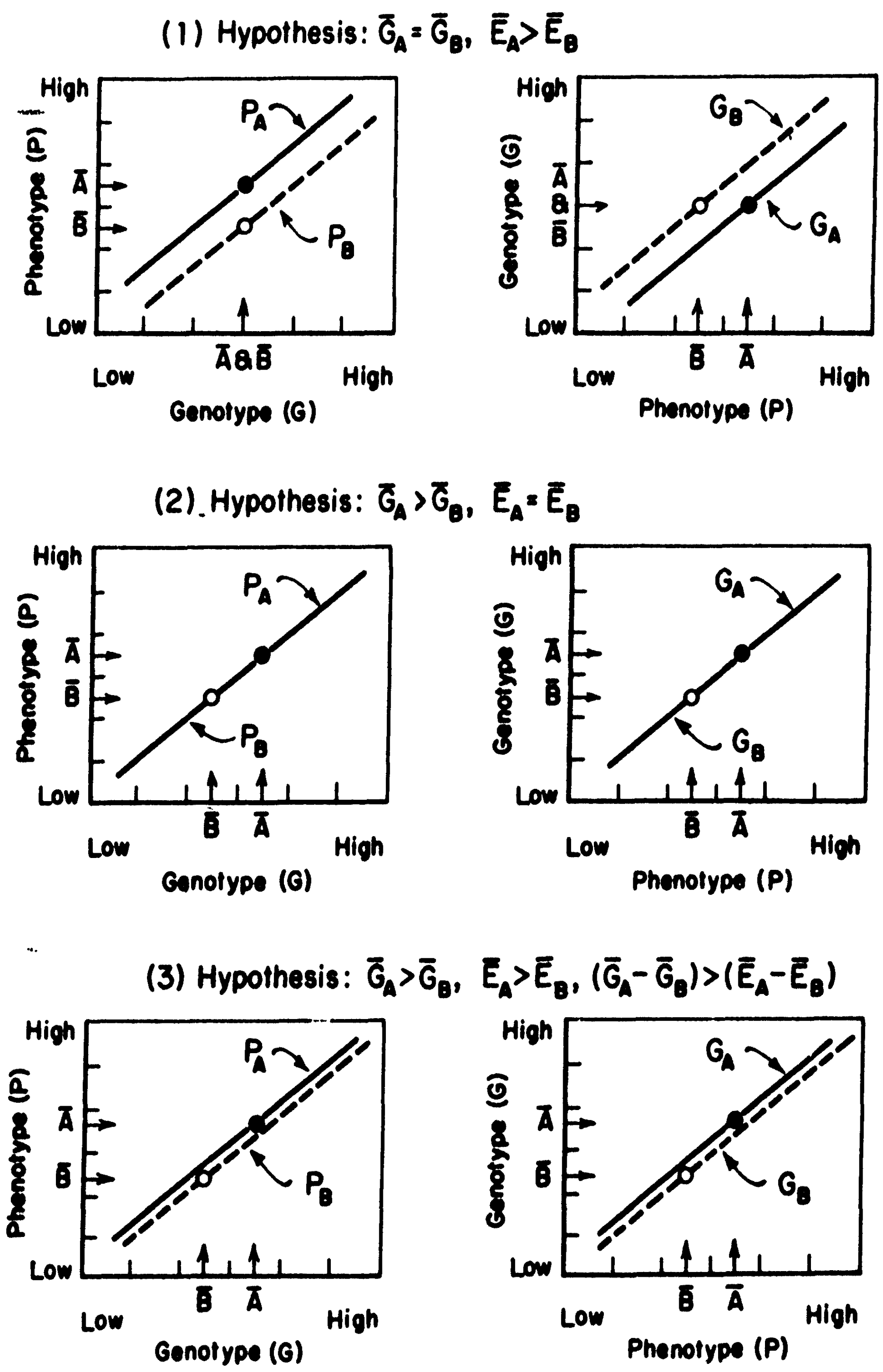


Figure 18. Models showing the predicted regression lines of phenotype on genotype (and vice versa) for two groups differing in mean phenotype under three different hpyotheses of the cause of the groups' phenotypic difference: (1) environmental, (2) genetic, (3) combined environmental and genetic.

tability of the phenotypic measure, $\underline{P}$, is .64, so the correlation between genotype and phenotype would be the square root of .64, or .80. Also assume that the means of the two groups, $\bar{A}$ and $\bar{B}$, differ on the phenotypic measure by 1 SD.

Hypothesis 1, then, is the environmental hypothesis. It states that the mean genotypes of the two groups are equal ($\bar{G}_A = \bar{G}_B$), and the average environment of group A is more favorable than that of group B ($\bar{E}_A > \bar{E}_B$). If this hypothesis is true, and if h^2 is .64 in each group, then the regression of $\underline{P}$ on $\underline{G}$ and of $\underline{G}$ on $\underline{P}$ for groups A and B should appear as shown in Figure 18 in the two graphs at the top. That is to say, for any value of $\underline{G}$, the value of $\underline{P}$ for group A will exceed that of group B by 1 SD. (The dots represent the bivariate means of groups A and B and the solid and dashed lines are the regression of $\underline{P}$ on $\underline{G}$ or $\underline{G}$ on $\underline{P}$.)

Hypothesis 2 is a strictly genetic hypothesis; the groups differ in genotype but not in environment ($\bar{G}_A > \bar{G}_B$ and $\bar{E}_A = \bar{E}_B$). Here we see that the regression of $\underline{P}$ on $\underline{G}$ (and $\underline{G}$ on $\underline{P}$) is the same line for both groups.

Hypothesis 3 is a combined genetic and environmental hypothesis, with two parts: (i) group A is more advantaged than group B both genetically and environmentally ($\bar{G}_A > \bar{G}_B$ and $\bar{E}_A > \bar{E}_B$), and (ii) the genetic difference is greater than the environmental difference, ($\bar{G}_A - \bar{G}_B > \bar{E}_A - \bar{E}_B$). Note that in this case the regression line P_A is above P_B, as in the top left graph (Hypothesis 1), but unlike Hypothesis 1, in Hypothesis 3 the regression line G_A remains above G_B.

Now, with the consequences that logically follow from these three clearly formulated hypotheses made explicit, as shown in the regression line of Figure 18, we can perform an empirical test of these hypotheses. Naturally, we can only crudely approximate the idealized hypothetical regressions shown in these graphs, since there are no perfectly culture-free tests,

i.e., tests with $h^2 = 1.00$. The best we can do at present is to use two tests which differ most conspicuously in culture-loading. (The most culture-loaded test corresponds to *P* in Figure 18 and the least culture-loaded test corresponds to *G*.) For this purpose we have chosen Raven's Matrices and the Peabody Picture Vocabulary Test (PPVT). We have already pointed out that the Raven is one of the most culture-reduced tests availabe. The PPVT provides a striking contrast. It is probably the most culture-loaded among all standardized IQ tests currently in use. The test consists of 150 plates each containing four pictures. The examiner says a word that labels one of the four pictures in each set and the testee is asked to point to the appropriate picture. The items increase in difficulty by increasing the rarity of the pictured objects and their corresponding verbal labels. Figure 19 shows the mean frequency of these words per every million words of printed English in American books, magazines, and papers. It can be seen that for both equivalent forms of the test (A and B), the commonness of the words decreases systematically from the first, easy items to the last, most difficult items. The PPVT pictures and labels are almost a parody of culture-biased tests: *e.g.*, *kangaroo*, *caboose*, *thermos*, *bronco*, *kayak*, *hassock*, *goblet*, *binocular*, *idol*, *observatory*, *oasis*, *walrus*, *canine*.

The Raven and PPVT were given individually to all white (*N* = 638), Negro (*N* = 381), and Mexican-American (*N* = 684) elementary school children in one small California school district.[83] The raw scores on both tests, within 6-month age intervals, were transformed to *z* scores, with mean = 0, *SD* = 1. The regression of Raven on PPVT and of PPVT on Raven was then plotted separately for each ethnic group. The regression lines are perfectly

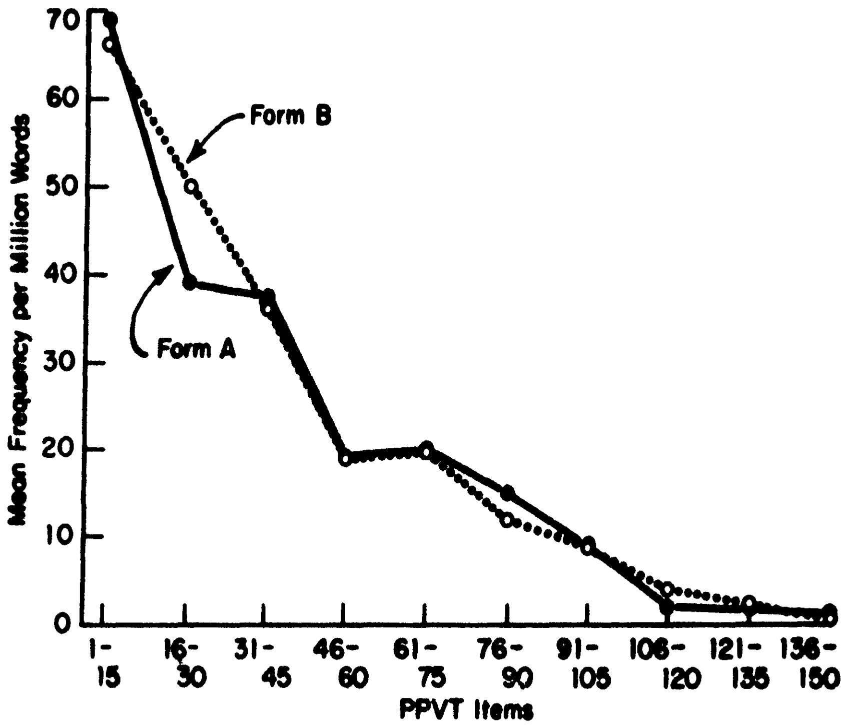


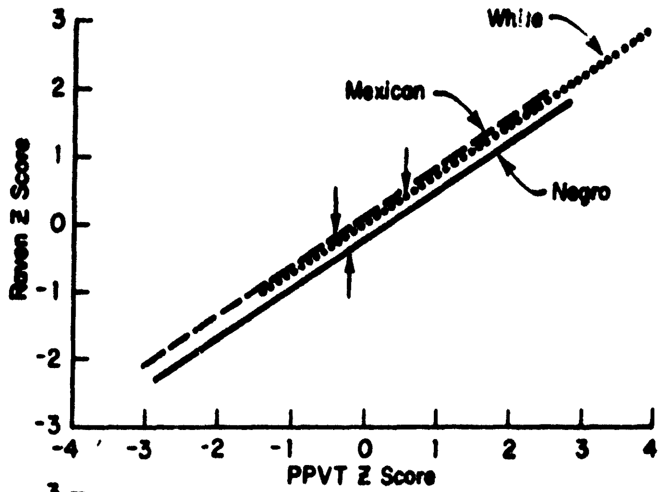
Figure 19. Mean Thorndike-Lorge word frequency of Peabody Picture Vocabulary Test items (for Forms A and B) as a function of item difficulty when items are ranked from 1 to 150 in P values, (percent passing) based on the normative population.

linear throughout the entire range of test scores in all three groups, as shown in Figure 20. The slopes of these regression lines of the three

groups do not differ significantly, but the intercepts differ significantly beyond the .001 level ($F = 52.38$, $df = 2,1658$). In short, the differences essential to our hypotheses are fully significant. So let us compare these empirical regression lines with the hypothesized ones in Figure 18. First, consider the white-Negro comparison (corresponding to hypothetical groups A and B). We see that the top half of Figure 20 corresponds to the left-hand graphs in Figure 18. Now we see that in both graphs of Figure 18 the white regression line is significantly above the Negro regression line. The only hypothesis to which this situation corresponds is Hypothesis 3 in Figure 18. Hypotheses 1 and 2 are both contradicted by the data.

Next, consider the white-Mexican comparison. Here we see that the Mexican regression line is <u>above</u> the white regression line for the regression of Raven on PPVT (upper graph in Figure 20), and the Mexican regression line is <u>below</u> the white regression line for the regression of PPVT on Raven (lower graph in Figure 20). This state of affairs is predicted only by Hypothesis 1. Thus we see that the results for the Negro-white comparison are predicted by one hypothesis (Hypothesis 3), and the results for the Mexican-white comparison are predicted by another, although both the Negro and Mexican groups are regarded as disadvantaged and score lower than whites on IQ and scholastic achievement tests. It is most interesting that each of the two sets of ethnic comparisons is consistent with a <u>different</u> hypothesis.

Finally, consider the Negro-Mexican comparison. For the regression of Raven on PPVT the Mexican regression line is above the Negro, but just the reverse is true for the regression of PPVT on Raven. This situation is not



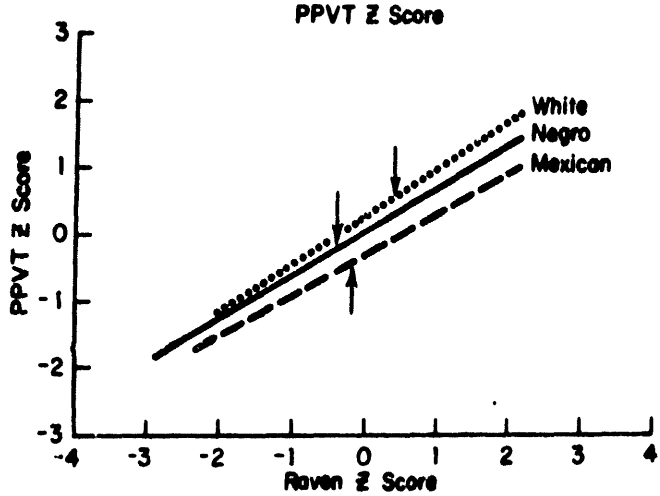


Figure 20. Regression of Raven's Matrices standardized scores (z) on Peabody Picture Vocabulary Test z scores (above), and regression of PPVT scores on Raven scores (below).

hypothesized in Figure 18, but it is what one would predict from the hypothesis $\bar{G}_A > \bar{G}_B$ and $\bar{E}_A < \bar{E}_B$, where A and B represent the Mexican and Negro groups, respectively. That is, the finding is consistent with the hypothesis that the Mexican group is genetically favored but environmentally or culturally disadvantaged relative to the Negro group. The results are well comprehended within the framework of these alternative hypotheses. Those who think in terms that are exclusively environmental, however, are invariably deeply puzzled by the results shown in Figure 20. If (in the lower graph) for any given score on the less culture-loaded test (Raven) whites get the highest score on the more culture-loaded test (PPVT) and Mexicans get the lowest, with Negroes intermediate, it seems to make perfectly good sense from the culture-bias or environmentalist hypothesis. But then when we look at the upper graph in Figure 20, we see that for any given score on the culture-loaded test, the Mexican gets the highest score on the culture-fair test, and this surely seems to make sense from the environmentalist standpoint. But, lo and behold, the Negro group's regression line does not come next -- instead it is well below the white group's regression line. In other words, if you match Negro, Mexican, and white children on the culture-loaded test, their scores on the more culture-fair test come out with Mexicans highest, Negroes lowest, and whites intermediate. This seems paradoxical to the environmentalist. It is predictable from the hypothesis formulated in Figure 18, which involves hypothesizing group differences in <u>both</u> genetic and environmental factors for explaining the Negro-white and Negro-Mexican differences. On the other hand, for these data at least, the hypothesis of only an environmental difference is compatible with the Mexican-white comparison. This methodology is presently being extended to other tests and other subpopulations. In terms of these formulations, it is already apparent from preliminary analyses that California Orientals bear a similar relationship to whites as the Mexicans bear to the Negroes, that

is, a higher average genotype and lower average environmental advantages.

<u>Piaget's Méthode Clinique</u>. My work has been severely criticized by one developmental psychologist on the grounds that my conceptions of intelligence and its measurement were based on standard psychometric tests, and he claims that "IQ tests are simply not adequate to measure processes of thinking" (Voyat, 1970, p. 161). He suggests that, instead of IQ tests, the theory and techniques of Jean Piaget, the noted Swiss child psychologist, should be applied to the problem of comparative mental development: "Piaget's approach not only allows an understanding of how intelligence functions, but describes it. Since the interest of Piaget's tests lies in describing the mechanism of thinking, they permit an individual, personalized appraisal of further potentialities independent of the culture" (p. 161). "In contrast, IQ tests, designed by whites for Western culture, have value limited to the culture within which they were designed" (p. 160). In view of these opinions, we must take a more detailed look at the relevance of Piaget's approach, which he calls the <u>méthode clinique</u>, to the study of children's mental development.

Briefly, Piaget views the mental development of the child as going through four main stages, which are invariant in sequence for all children; (1) The <u>sensorimotor</u> stage (onset from birth to about 1 year) is the first phase of intellectual development, in which knowledge and thought are intimately tied to the content of specific sensory input or motoric activity of the child; it includes conditioning, stimulus-response learning, reward learning, perceptual recognition, and associative or rote learning and memory. (2) The <u>preoperational</u> stage (onset ages 1 to 2 years) is a transitional period between the sensorimotor stage and the next stage and is mainly characterized by symbolic play and cognitive egocentrism, <u>i.e.</u>, the child in this stage can view objects and relationships only in terms of his

own relation to them. (3) Concrete operations (onset 6 to 7 years) is the first stage of what Piaget calls operational thinking, which characterizes his view of intelligence. It involves the capacity for performing mental operations on concrete objects, such as numeration, seriation, and classification or other forms of grouping, and the ability to conceive the invariant structure of classes, relations, and numbers. (4) Formal operations (onset 11 to 13 years) is the final level of operational thinking, manifested in logical reasoning (not dependent upon the manipulation of concrete objects), propositional thinking, combinatorial and inferential thinking which involve using hypothetical possibilities, abstractions, and imaginary conditions as well as the mental manipulations of symbols for real or experiential knowledge.

Piaget has devised a large number of ingenious "test" or clinical-type procedures for assessing the child's mental development as he moves through these stages, each of which has finer gradations or substages marking the course of cognitive development. Most of the techniques have concentrated on the assessment of concrete operations, for this is the first stage of operational logical thinking which, in Piaget's view, is the beginning of mature intelligence and most characterizes human intelligence. The child's capacity to grasp and utilize the concepts of conservation of number, weight, and volume, in that order, marks the development of operational thinking. The 7 or 8 year old child who is well along in concrete operations, for example, is capable of conserving volume; that is, the quantity or volume of a ball of clay or a jar of liquid is perceived as invariant regardless of its changing shape (a round ball of clay or the same ball of clay flattened out like a pancake) or the variety of differently shaped flasks in which the liquid can be poured (low, flat bowl or tall, thin cylinder). The preoperational child cannot maintain this invariance; to him, when a round ball of clay is flattened out and made to look "bigger," he actually belives the

quantity of clay has been increased, and similarly when he sees liquid poured from a shallow, broad bowl into a tall, slender flask. There are many ways that the concept of conservation shows up: in number, length, area, time, weight, volume, and so on. Piaget has invented means for assessing children's conservation concepts in all these forms, along with many other tests and procedures for studying the sequence of mental development throughout each of its main stages.

Now, what have child psychologists learned from the application of Piaget's tests that is relevant to Voyat's commentary?

First, Voyat is probably correct in his opinion that the Piagetian tests are less culture-bound than conventional IQ tests. For one thing, some groups reared under environmental conditions which are extremely different from those of Western culture have been found to show not only the same sequence of development through Piaget's stages, but are even somewhat more accelerated in this development than white middle-class children. Again, Arctic Eskimos were found to excel over white urban Canadian children in the Piagetian tests, and Canadian Indians do almost as well as the Eskimos (MacArthur, 1968, p. 48; Vernon, 1965b). Obviously it is not necessary to have lived in a Western or middle-class culture in order to perform up to Western middle-class levels on Piagetian tests. In fact, numerous studies have shown that formal schooling has no effect on the age of achieving the various component structures and skills that comprise "concrete operations" (Kohlberg, 1968).

In rank-ordering children of the same chronological age in terms of their rate of mental development, the Piagetian tests are not very different from other culture-reduced tests. Vernon (1965b) factor analyzed a large number of Piagetian tests along with conventional psychometric measures of intelligence and found that the Piagetian tests were heavily loaded on *g*,

the general factor common to all intelligence tests. In fact, the Piagetian tests measured little else than *g*; the non-*g* variance seems to be task-specific, *i.e.*, it has nothing in common with other Piagetian tests or with conventional IQ tests. Tuddenham (1970) gave a battery of Piagetian tests, along with Raven's Matrices and the Peabody Picture Vocabulary Test (PPVT), to a large number of elementary school children, and concluded: ". . . the Raven has the higher correlations, ranging from 0.24 to 0.50, as compared with Peabody values of 0.13 to 0.37 for a similar though not identical set of Piagetian items" (p. 68). These are relatively high values for single item correlations within a restricted age range. Tuddenham notes that "Correlations with Piaget item composites of six and eight items respectively are 0.60 for the Raven *vs.* 0.21 for the Peabody." These are the kinds of correlations one should expect if Piaget's tests are culture-reduced, since among psychometric tests, the Peabody and the Raven are probably further apart than any other tests on the continuum going from "culture-loaded" to "culture-free."

Do the Piagetian tests have high heritability? It would be most surprising if they did not, in view of what has just been said, although there have not yet been any heritability studies of these tests. One impressive study, however, strongly supports the idea that Piagetian tests are highly sensitive indicators of genetic factors in mental development. DeLemos (1969) administered a battery of Piagetian tests to Australian aborigines, ages 8 to 15 years. The aboriginal children were remarkedly retarded as compared with European and American norms. The majority of adolescents were still not up to the level attained by the average European 7-year-old. Even the majority of aboriginal adults do not reach the level of concrete operations represented by the conservation of quantity and volume, although there a few exceptions. In the course of this study, DeLemos compared the

Piagetian test performances of full-blooded aborigines with those who were part aboriginal and part Caucasian. The children's ancestry was known from records kept by the mission in charge of the territory inhabited by these aborigines. The Caucasian ancestors were "casual," probably being immigrant laborers and sailors, and never lived among the tribe. "Among the children classified as part-Aborigines the degree of European ancestry was small, the majority being classified as 7/8th Aboriginal [the equivalent of 1 Caucasian great-grand parent]. The European ancestry was therefore several generations removed from the present group. There were no apparent differences in the present environment of the part-Aboriginal and full-Aboriginal children. . . . Part Aborigines and full-Aborigines formed a single integrated community, and the children were brought up under the same mission conditions and attended the same school" (DeLemos, 1968, p.257). What DeLemos found was that the part-aboriginal children were markedly advanced in the Piagetian measures as compared with the full-aborigines of the same ages. The differences were not small and did not depend upon large samples for their high level of statistical significance. They are remarkably large differences, beyond anything that has ever been produced by direct training on Piagetian tasks and concepts. The results for the six types of Piagetian conservation tests used by DeLemos are shown in Table 4.

The results appear almost as if the admixture of Caucasian genes, even so few as 1/8th, introduces mental structures otherwise lacking, that permits the individual to reach higher levels of mental development than normally occurs in the majority of full-Aboriginals. Commenting on this striking finding, DeLemos writes: "The significance of our results lies in the fact

Table 4

Comparison of the Number of Part Aboriginal and Full Aboriginal Children Showing Conservation*

Test	Full Abor. $N = 38$	Part Abor. $N = 34$	χ^2	p
Quantity	4	18	15.21	$< .001$
Weight	16	25	7.23	$< .01$
Volume	2	8	3.59	$.05 < p < .10$
Length	12	20	5.37	$< .05$
Area	3	10	4.23	$< .05$
Number	3	9	3.22	$.05 < p < .10$
Total	40	90	36.14	$< .001$

*From DeLemos (1968)

that in this case there were no apparent differences in the environments of the two groups. Both formed an integral part of the same community, being closely related by family and kinship ties, and living under the same mission conditions. The differences cannot therefore be attributed to environmental factors. . . .It would therefore seem reasonable to attribute the significant differences between the part and the full Aborigines in this study to genetic differences between Aborigines and Europeans, resulting in the part Aboriginal children having a higher probability of inheriting a higher intellectual potential" (p. 268).

Finally, what do the Piagetian tests reveal about the cognitive development of American Negro children? Read Tuddenham (1970) carried out the major study, giving a battery of ten Piagetian tests to some 500 white, Negro, and Oriental children in grades 1 to 3 in three California communities. Negroes did less well than whites on every item. The average percentage of children possessing the concept tested by the particular items was 32.6 for whites <u>vs</u>. 15.9 for Negroes. If the items are combined and the scores treated in the manner of conventional psychometric tests, the white Negro difference would amount to approximately 0.7σ (equivalent to 10 or 11 IQ points). Oriental children, on the other hand, were more advanced than white children on 7 of the 10 items. The Piagetian scale also correlates substantially with SES as indexed by father's occupation, even though, as Tuddenham notes, "these items tend to involve reasoning about matters universally available to observation, <u>e.g.</u>, the horiziontality of water levels. It is hard to see how social advantage could be a very large factor in success on some of these items. The genetic selection implicit in occupational level may well have more to do with it" (p. 65).[84]

Gaudia (in press) administered a series of Piagetian conservation tasks to 126 low SES American Indian, Negro, and white children in grades 1 to 3. Overall, these groups, all being of very low SES, averaged about 1 year behind the age norms on these tests based on samples of the general population. But the Negro children in this study were significantly ($p < .001$) delayed in the acquisition of conservation (of area, number, quantity, weight, and mass) as compared with the low-SES Indians and whites, who did not differ significantly. The racial disparity was greatest in the older age groups. Expressed as a percentage of the highest possible conservation score, the means of the three age-matched ethnic groups are: white = 51, Indian = 51, Negro = 30.

How much does specific training in attention and classification raise children's performance in these Piagetian tests? To find out, Sigel and Olmsted (1970) gave one month of training on certain skills and concepts intended to promote cognitive development to Negro children enrolled in a Headstart program. A year after the training, these children were compared with a matched control group on five Piaget tests of logical operations (multiple classification, multiple seriation, and reversibility) and conservation (number and quantity). The training is reported to have had no significant effect on performance in any of these tests. The authors state, "These results cannot be attributed to a lack of understanding of concepts like more, same, or less, since tests for this were administered. All children passed this test. . . . It is important to point out that the difficulties these first-grade children have (both at the beginning and at the end of first grade) in not being able to conserve number and/or mass, reveals the seriousness of their cognitive deficit, especially if the criterion used is our data from middle-class whites. It is worth pointing out that among 5-year old white middle-class children, conservation of

number and mass are soluble. Of the 75 (white) children tested . . . approximately 50% could conserve in these areas without training and after a nine-session training program, 68% of the previous non-conservers could then conserve" (Sigel & Olmsted, 1970, p. 328). By contrast, of the Negro Headstart children who received the one month of training, ". . . 81.3% could not conserve numbers and 93.2% could not conserve quantity either before or after training" (p. 328).[85]

Sensorimotor Differences

There has never been any real disagreement about genetically determined physical differences -- biochemical, physiological, and anatomical -- between racial groups. It is in the realm of intellectual functions that we see so much *a priori* resistance to rejection of the null hypothesis concerning genetic racial differences. But actually there is a continuum between the physical and the intellectual; there are no discernible discontinuities; and racial differences are found at all points along the continuum going from strictly physical characteristics to behavioral characteristics including those processes we identify as mental ability. So the point at which one draws the line of resistance to entertaining a genetic hypothesis of racial differences must be completely arbitrary and without any logical or scientific basis.

Sensory capacities are intimately related to physical structures and processes and are undoubtedly conditioned by genetic factors. And we find marked racial differences in certain sensory capacities. (An excellent detailed review of much of this evidence is found in Spuhler & Lindzey, 1967.) For example, the ability to taste the synthetic chemical substance phenylthiocarbamide (PTC) is known to be genetically determined, probably by a single gene. There are striking race and subpopulation differences in the frequencies of tasters and non-tasters of PTC, going from 0 to 57

percent in the various populations which have been studied (see Spuhler & Lindzey, 1967, pp. 381-384). There are marked racial differences in the incidence of various types of sex-linked color blindness which are completely genetic; and there are differences in ability for color discrimination. Negroes have better visual acuity than whites; only 65% of whites in the armed forces pre-induction examination have 20/20 vision, as compared with 82% of Negroes (Dreger & Miller, 1968, p. 7). Negroes also show better dark adaptation than whites. Negroes show a greater galvanic skin response than whites, and they perceive radiant heat at a lower threshold (Dreger & Miller, 1960, p. 364).

Moving along the behavioral continuum from sheer sensory to more perceptual processes, population differences have been found in degree of susceptability to various optical illusions (see Spuhler & Lindzey, 1967; Dreger & Miller, 1960, 1968). It is virtually impossible to explain some of these illusions in terms of cultural or experiential differences. Groups whose visual experiences are highly similar may differ greatly in susceptibility to a particular perceptual illusion and groups whose environments differ markedly may show no differences in the illusion. One illusion (the black-white radiation size illusion) which is explainable in physiological terms on the condition that the illusion stimuli are presented in black and white but not in response to variously colored illusion stimuli. It is significant, therefore, that there are racial differences in susceptibility to this size illusion when it is presented in black and white but not in response to variously colored illusion stimuli.

<u>Speed of Visual Information Processing</u>. As we move along the continuum to speed of visual information processing, we come somewhat closer to mental abilities. Intelligence, in fact, is sometimes defined as

information processing capacity. So it should be interesting to look at the simplest form of visual information processing, which comes very close to being almost a physiological measure of a basic mental capacity.

Information processing capacity shows up in a most fundamental form in a phenomenon technically known as meta-contrast or "masking." If a visual stimulus is presented to the observer's view for a standard duration, followed by a "blank" interval, and then by a second stimulus (equal to or greater in area than the first) of standard duration, the observer either will or will not be able to name the first stimulus (e.g., a letter of the alphabet) depending upon the duration of the "blank" period between the first and second stimuli (called the "test signal" and the "masking stimulus," respectively).[86] Studies have revealed highly reliable individual differences in the shortness of duration of the "blank" period or interstimulus interval that observers can tolerate without "losing" the test signal. If the blank interval is too short, the test signal is literally wiped out and the observer's guess as to what it was is no better than chance. The duration of the shortest interval at which the observer can identify the test signal has been called "information processing rate."

A meticulous study by Bosco (1970) has shown marked subpopulation differences in information processing rate. He compared children in first, third, and sixth grades in urban schools attended by low SES and middle SES children. Race and SES are confounded in this study (Low SES: 28 whites, 62 Negro; Middle SES: 88 whites, 2 Negro), so we shall refer to the contrasted groups only as low SES and middle SES. The four test signals were very carefully selected so as to eliminate experiential differences. They consisted of circle, square, triangle, and five-pointed star. Bosco comments: "The stimuli which were used in this study are so pervasive as to rule out any possibility of them not being present within the low SES

environment. None of the children in the study had difficulty identifying the four stimuli during the preliminary part of the testing. Even the disadvantaged first graders responded correctly and promptly" (p. 61). Bosco also reported that observations during the testing did not lead him to suspect motivational differences between the two groups. With four stimuli each having equal probability of occurrence, there were only "bits"[87] of information transmitted in this procedure, thus making it a very rudimentary but clear-cut measure of information processing capacity.

Bosco found large significant ($p < .01$) differences in information processing rate between the low and middle SES groups. The mean SES difference was greater than the mean grade difference. First-grade middle SES children had a slightly faster processing rate than low SES children in the sixth-grade. As one would expect for a processing task involving only 2 bits of information, the group differences (as well as individual differences) decrease with age. At first grade, low SES children require more than twice as much visual processing time as needed by middle SES children. At sixth grade, the low SES children used about 30 percent more time. Bosco found low correlations (around .20) with various scholastic achievement measures. (Unfortunately, these correlations were obtained only in the sixth-grade, in which there was relatively little variance in visual processing rate.) Bosco comments, "The more a variable assesses a basic cognitive variable, the less likely we ought to expect relationship to school success. . . As it is, there is good reason to think that school achievement is a result of a host of variables in addition to cognitive variables," (p. 51). It seems likely, however, that Bosco underestimates the relevance of his measure of scholastic performance. This could be tested by determining correlations at the first grade, and by increasing the information load of the task at sixth grade in order to yield sufficient variance to permit significant

correlations to show up.

Reaction Time. Reaction time (RT) to a stimulus situation increases as the amount of information transmitted by the stimulus increases. RT increases as a linear function of "bits" of information. Also, it has been found that the slope of this function is negatively correlated with IQ (Eysenck, 1967). A description of one experimental procedure for demonstrating this will help to make it clear. The subject sits in front of a panel on which there is a single light bulb; directly beneath the bulb is a pushbutton. When the light flashes "on," the subject pushes the buttton to turn the light "off." In this condition, the subject's response time is a measure of simple RT. There is zero information conveyed when there is only one light/button combination. But the subject is required to respond to an increasing number of light/button combinations, simply by having one light go on among an increasing number of potential alternatives. This is called "choice RT." The amount of information conveyed increases logarithmically as the number of lights. Now, when zero information is conveyed, there is no correlation between RT and IQ. With an increasing number of choices, RT correlates increasingly with IQ. This relationship has been demonstrated by Roth (1964). And Fox and Taylor (1967) compared two groups of army recruits on simple RT and choice RT. One group (Low AFQT) was selected from recruits having scores between 10 and 21 on the Armed Forces Qualification Test (a composite measure of general intelligence and basic scholastic attainments); the other group (High AFQT) were recruits with scores from 90 to 99. The groups differed significantly in choice RT, but not in simple RT. (A more detailed description of the apparatus, procedure and results of this experiment is presented in Jensen, 1970d, pp. 149-151.)

Noble (1969) gave a 4-choice reaction time test to groups of rural Georgia white and Negro children (N = 106 in each group) matched for age

and sex. Each child was given 160 standard trials. The results, plotted in terms of mean response speed (the reciprocal of RT) for blocks of 20 trials, are shown in Figure 21. The overall white-Negro difference is

significant ($p < .01$). Note that response speed increases with practice, but soon levels off in both groups. The first trials show no Negro-white difference, and the mean difference in the first block of 20 trials is small as compared with later blocks. If motivational and attitudinal factors were acting to depress the performance of the Negro children, it is hard to see why they should have differed so little at the beginning of practice. Increased practice tends to increase and stabilize the magnitude of the difference between the groups.

Motor Skills Learning. Noble (1968, pp. 230-231; 1969) has reported an exemplary study of motor skill learning in Negro and white rural Georgia school children, ages 9 to 12. From a pool of 500 subjects, all right-handed, 152 were selected so as to form four groups, each with 38 subjects (two Negro and two white groups) matched for age and sex.

The task was pursuit rotor learning. The pursuit rotor is the most widely used instrument in laboratory studies of human motor learning. Hundreds of experiments have been performed with the pursuit rotor and much more is known about experimental parameters of performance on this motor skill than on any other. It is a "tracking" skill. The apparatus consists of a disc about the size of a phonograph turntable which rotates at a given speed (usually 60 r.p.m.). The disc is made of a smooth non-conductor such as bakelite; flush with the disc's surface and about halfway between the center and the edge is a "target" -- a small silver metal disc, usually

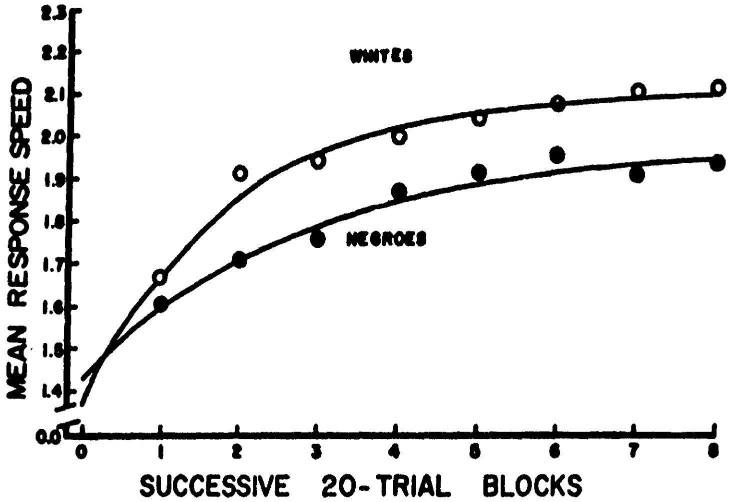


Figure 21. Mean response speed in successive 20-trial blocks on choice reaction time. Each curve based on 106 children. (From Noble, 1969)

about 3/4 inch in diameter. As the turntable rotates, the subject's task is to learn to keep a metal stylus on the target; the stylus is hinged to its handle in such a way that no pressure can be exerted by the subject as a means of keeping the stylus "on target." When "on target" the stylus completes an electrical circuit, activating a timing device which records the percentage of each consecutive 10-second period that the stylus is on target. Learning, that is, improvement of the skill with practice, is reflected in the increasing average percentage of time on target when the course of practice is divided into a number of periods of equal duration.

Pefore Noble performed his experiment, a number of relevant factors were already known about pursuit rotor learning. For one thing, this form of learning has not been found to be sensitive to examiner effects; that is, the sex, age, race, and attitude of the experimenter do not significantly affect the subject's performance. Even so, Noble took precautions in his study. He used both male and female Negro and white experimenters, counter-balanced for all groups in the experiment. (He found no statistically significant effects on tracking performance attributable to sex or race of the examiner.) Also, he minimzed any possible experimenter influence by leaving the child alone in the testing room after the instructions were given. (Instructions were given largely by means of demonstration by the experimenter.) As a further check, he recorded the subject's pulse rate just before and after the learning period, on the assumption that if there were any differences between the groups it would show up in the pulse rate, which is a sensitive indicator of anxiety. There was no race difference and no pre-post test difference in pulse rate. The children were not anxious but actually enjoyed the task and the fun of taking turns and getting out of their regular class activites to participate in the experiment. Also, there

was no prior evidence that pursuit rotor learning has any appreciable correlation with intelligence. In a group of 186 boys, for example, McNemar (1933) found a correlation of only .17 between tracking ability and IQ. Obviously, not all kinds of learning ability are as highly related to IQ as is scholastic learning. Finally, it was known that pursuit rotor learning has very high heritability, almost as high as the heritability of height. NcNemar (1933) obtained correlations of .95 and .51, respectively, for MZ and DZ twins. Using the simplest formula for estimating heritability ($\underline{h}^2 = 2(\underline{r}_{mz} - \underline{r}_{dz})$, which assumes no assortative mating for pursuit rotor ability, the value of $\underline{h}^2$ obtained from McNemar's data is .88. Furthermore, Vandenberg (1967) reports that heritability is much higher for pursuit rotor learning with the right hand (or preferred hand) than with the left. In other words, the tracking task can serve either as a test having very high heritability or as a test having low heritability, depending on whether the subject is required to use his preferred or his non-preferred hand.

With this background in mind, Noble had half of each racial group (all were right-handed) perform with their <u>right</u> hand and half of them with their <u>left</u> hand. The results are shown in Figure 22. The white

subject's average left-hand performance was slightly better than the Negro's performance with the right hand. Also, the race difference is much greater for the right-hand performance, with its higher heritability.

So striking and interesting were these results that Noble replicated and extended the study on a new group of 268 subjects, and obtained essentially the same results, significant beyond the .001 level: "Whites not only performed at a generally higher level of proficiency than Negroes but also

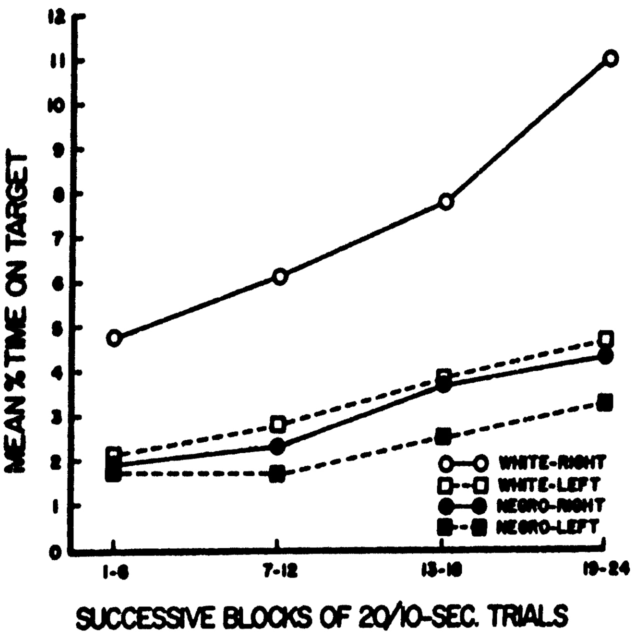


Figure 22. Mean percentage of time on target in successive blocks of 20-second work and 10-second rest trails on the pursuit rotar, for white and Negro children, practicing with either the Right or the Left hand. There are 38 subjects in each condition. (From Noble, 1969)

were gaining at a faster rate. Even after fifty practice trials conducted under rigorously controlled conditions, the average Negro right-hand ability was still below the average white left-hand ability" (Noble, 1969, pp. 22-23).

Noble then went a step further. He divided the Negro group into two groups which we shall call blacks and mulattoes. He used several genetically independent (but phenotypically correlated) objectively measured physical criteria for this classification, and showed that the groups differed significantly on each one: skin pigmentation, nasal width, lip thickness, hair texture, eye color, jaw formation, interpupillary distance, and ability to taste phenylthiocarbamide. The subjects thus classified into three groups showed significantly different mean pursuit rotor scores in the order: whites > mulattoes > blacks. The mean percentage of time on target for the three groups were 4.6%, 2.6% and 2.1%, respectively (Noble, 1968, p. 231). Noble believed that strictly environmental interpretations of these results in terms of socioeconomic and cultural differences would find little evidential support. "On the contrary," he writes, "there were numerous observations to suggest that a large part of the interracial behavioral variance is genetically determined" (1969, p. 27). But Noble concludes: "Our data on the learning of psychomotor skills pertain to operationally defined concepts and to statistical aggregates of subjects; nomothetic laws imply nothing about (1) the human worth of a particular individual or (2) their civil liberties. Whether all the living races of mankind are equal in their innate (genotypic) biological potentialities for cultural and scientific achievements in modern civilization is a matter that cannot be determined by single experiments of deliberately limited scope. At the same time, I hold that systematic, theory-oriented, fundamental research on behavior offers our best hope of solving the vexing interracial problems that confront us today" (p. 29).

Physical Environment and Mental Development

Nutrition. In recent years poor nutrition has been frequently mentioned in the literature as if it were established as a major environmental cause of social class and racial differences in intelligence and scholastic performance. Hence we must briefly review what is actually known at present about the effects of nutrition on mental development and weigh its relevance to the issues under discussion in this paper.

First of all, the evidence from experimental studies of severe nutritional deprivation in animals leaves no doubt whatsoever that brain functions and their behavioral correlates are influenced by nutritional factors, particularly protein deficiency (Dobbing, 1970; Eichenwald & Fry, 1969; Scrimshaw & Gordon, 1968; Winick, 1970). The authors of animal studies of nutritional deprivation showing effects on brain and behavior usually point out that the deprivation is severe, that is has generalized effects on the growth and functioning of other systems, and that larger and more irreversible effects are produced by nutritional deprivation during the periods of most rapid growth of the central nervous system.

The evidence with respect to the psychological effects of malnutrition in humans is much more difficult to evaluate, since it cannot be based on experimental studies but must rely on the occurrence of nutritional deprivation in "natural" settings. This invariably means that the effects of malnutrition are confounded with a host of other unfavorable factors so frequently associated with poverty and with poor mental development, such as prematurity, low birth weight, poor health care, high incidence of infectious diseases, child neglect, and so on. Also, as Platt (1968, p. 241) points out: "The effects of maternal undernutrition and those of genetic factors are difficult to separate in disadvantaged populations. Since consequences may be similar and are exhibited even in fetal death, there is often no way to separate

these factors in the individual case. Human population and biostatistical studies must be conducted in order for scientists to understand the subtle interactions of genetic endowment and nutrition within, and between, the larger genetic pools of any given geographical or socioeconomic group." Two or three studies of human malnutrition, however, have involved more or less adequate controls and have yielded sufficiently clear-cut results to warrant the conclusions of one of the leading researchers in this field: "There can be absolutely no question about the association of significant degrees of malnutrition during the early years of life and concurrent as well as later manifestations of intellectual impairment" (Birch, 1968, p. 57).

I have found a total of only 13 published studies of the effects of nutrition on mental development. Eleven of these are well summarized by Stein and Kassab, 1970, who do not include a study conducted in Peru (Pollitt & Granoff, 1967) or the one published United States study (Harrell, Woodyard, & Gates, 1955). It is significant that all but one of the studies showing any mental effects of nutritional deficiency were conducted outside the United States in those parts of Africa, Asia, and Central and South American which suffer the most extreme poverty and protein-calorie deficiency.[88] It is interesting that even in these localities the degree of malnutrition sufficient to depress mental development is not found generally in any appreciable segment of any population; these malnourished cases must be sought out in specific families, and even then usually not all the children in the same family will show signs of malnutrition. The impression that many persons seem to have gained from the popular press, that there are hordes of malnourished children who suffer mental retardation therefrom, is quite at odds with the actual picture given by the total body of scientific literature on this subject. Children in whom mental effects of poor nutrition can be demonstrated have seemed almost as hard for researchers to find as

identical twins reared apart. The total number reported in the literature is fewer than a thousand, and in only a fraction of these have psychological effects been adequately demonstrated. The problem, of course, is that malnutrition is most often found in families in which frequently other factors, genetic and environmental, that cause mental retardation are also operative. The mean IQ of children in the general population, or even in that segment of the population in which the undernourished children were found, cannot serve as a proper comparison group because of this close linkage of malnutrition to other adverse familial factors. Those studies, such as Cravioto's (1968), which have used adequately nourished siblings as controls against which to measure the effects of malnutrition in the affected siblings are the most adequate, and their results leave little doubt that severe protein deficiency in the final months of fetal life and in infancy can depress important cognitive functions that emerge later in childhood. The statistical significance of this finding is not in question, but the magnitude of the effect is difficult to evaluate in terms of any familiar scales of mental measurement. The most adequate studies, carried on in Mexico and Guatemala by Cravioto and his colleagues, make use of special tests measuring "intersensory integration" (i.e., cross-modal transfer), and from the information given it is impossible to determine how much the significant differences between the nutritionally deprived and their controls would amount to on an IQ scale. I believe, however, that the measures of intersensory integration used by Cravioto et al. get at the g factor of intelligence at a very basic level (Cravioto & Delicardie, 1970). Thus, although we may conclude that there are significant effects of early malnutrition on cognitive development, little if anything can be said at present about the magnitude of these effects relative to the magnitudes of the social class and racial differences existing in the United States. One study using conventional individual

intelligence tests found an average of 20 points IQ deficit in a group of 21 severely malnourished African children as compared with an adequately nourished control group (which was more advantaged in many other respects as well). The degree of chronic malnutrition in these cases was so severe as to actually stunt physical growth and to cause abnormal brain waves; all the children had come to the attention of public health agencies, so their condition could not be regarded as typical even for the slums of South Africa (Stoch & Smythe, 1963).

A number of findings and conclusions are repeated in many of the studies. The degree of malnutrition associated with cognitive deficit is usually severe; in one study in Chile, for example, children during the first year of life were not breast fed but subsisted entirely on a diet of flour and water (Stein & Kassab, 1970, p. 95). The most severe retardation is associated with malnutrition in the last three months of gestation and in the first several months after birth, which is the most active period of brain growth. Retardation is less demonstrable among children who are malnourished after the second year of life, and there is evidence that older children and adults suffer no permanent psychological effects from even severe malnutrition, as existed in concentration camps near the end of World War II. Malnutrition is less often found where children are breast fed in the first year of life; artificial low-protein foods and "empty calories" in high carbohydrate diets are more often associated with poor nutrition. Early malnutrition hinders general growth and therefore causes an increased correlation between various physical indices and measures of intelligence.[89] Winick (1970) reported that at 2 1/2 to 5 years of age 70 percent of malnourished children had head circumferences below the 10th percentile -- a very skewed distribution indeed -- as compared with control children, whose head circumferences showed a normal distribution. Among malnourished children there is a significant

77-943 O - 72 - 26

correlation between head circumference and IQ, but no significant relationtionship was found in control children whose head size was within normal limits (Stein & Kassab, 1970, p. 101). Similar effects are found for height; malnutrition, particularly protein deficiency, retards the rate of ossification of cartilage in the first months of life (Platt, 1968, p. 243). Malnutrition also retards early motor development. In every study in which infant development tests, such as the Gesell scale, have been used, they show that the malnourished children score below par. Early malnutrition makes for greater inter-sibling differences; siblings within the same family are not equally affected, but in families in which malnutrition is found, there are significantly larger differences between the sibilings as compared with adequately nourished families (Cravioto & Delicardie, 1970). There is also evidence that protein deficiency impairs memory as well as other cognitive functions (Cravioto & Delicardie, 1970).

When a high percentage of low IQs are found among groups of children who themselves have shown no evidence of poor nutrition, it is hypothesized by some investigators that the lower IQs are a result, at least in part, of the children's mother's, or even grandmother's, having suffered from poor nutrition. There is some evidence for the intergenerational effects of malnutrition (and other environmental factors) on behavior in rats and dogs, but not in humans.[90] Stein and Kassab (1970, p. 109) summarize the present state of knowledge on this point: "There are no studies in human societies which can be held to support a cumulative generational effect of dietary restriction. Certainly any such effect was not sufficiently widespread, after countless generations of rural poverty, to prevent the emergence during the past century of the technological societies of Europe and North America."

Since all the studies mentioned so far are based outside the United States, we should look in more detail at the one published study conducted in the United States relating IQ to nutrition, and at another more recent study, as yet unpublished. The first study, by Harrell, Woodyard, and Gates (1955) was carried out in the Cumberland Mountains of Kentucky and in Norfolk, Virginia. The Kentucky subjects were "poor whites" living in what the authors describe as "deplorably low" economic conditions. The Norfolk subjects all were mothers on welfare, chosen for their low income status; 80 percent were Negro. There were 1,200 mothers in each group. These women early in pregnancy were contacted through public maternity clinics and given a variety of dietary supplements (one group got vitamins; another group got "polynutrients," and the control group got a placebo; i.e., a non-nutritive substance). These dietary supplements were taken throughout pregnancy. The children born to these mothers were given two forms of the Stanford-Binet Intelligence Test, at ages 3 and 4. There were 1,414 children tested in all. The Kentucky children showed no significant effects. (For both tests $\underline{F} < 1$.) The Norfolk children did show significant effects, however. At age 3 the vitamin and polynutrient groups averaged 2.5 to 5 IQ points higher than the placebo group. At 4 years of age the average gain over the placebo group was 5.2 IQ points in the vitamin group and 8.1 points in the polynutrient group. However, for both of these groups and the placebo group there was a significant ($\underline{P} < .001$) decline of 3.04 IQ points between ages 3 and 4. Thus, while the dietary supplements did raise IQ several points over the placebo group, they did not prevent the lowering of IQ between ages 3 and 4. This rapid decline within a one-year period, in addition to the fact that IQ at age four accounts for something less than 50 percent of the IQ variance in late adolescence, makes this study inconclusive as to whether any lasting effects on IQ were derived from

the dietary supplements during pregnancy. The IQs of the children at ages 3 and 4 were within the typical range for this population, and the decline in IQ from 3 to 4 is also typical; studies of similar groups have found average declines of about 10 IQ points between 3 and 6 years of age (Shuey, 1966, pp. 6-31).

The second study takes a still different approach, which consisted not of looking for children showing malnutrition and determining their psychological characteristics, but rather finding children in the poorest families in the poorest slums of a large Southern city, Nashville, Tennessee, (Carter, Gilman, Vanderzwaag, & Massey, 1971). The investigators visited community agencies to find out the location of poverty areas and to identify poverty families. These areas were then explored by car, followed by house-to-house canvassing by a social worker to find the most impoverished families with children of certain ages. The groups finally selected came from two housing projects on the East side of Nashville. The criteria for selection included: mother under 35 years of age, the target child should not be farther along in the family than the third child, and younger siblings should be present. Since the medical, nutritional, and psychological assessments were intended to be extremely thorough and elaborate, only 19 families were selected, 10 Negro and 9 white. The target children (singled out for special intensive study and enrollment in an experimental preschool program) were between the ages of 3 years 8 months and 4 years 8 months. The authors describe in general terms the typical backgrounds of the white and Negro families from which their samples were drawn: "The typical family of a white child . . . is likely to be one in which the natural father is present in the home at least 50 percent of the time. He is usually an unskilled laborer or perhaps disabled. The average annual income is below the OEO Poverty Guidelines. Half of the mothers were on Welfare or Aid to Dependent

Children Programs." (About 40 percent of the white mothers had completed high school) "The typical black family . . . is likely to be one in which the natural father is not at home. If he is, he is usually employed in maintenance work, in the military service, or is a trainee in some OEO program designed to find jobs for the hard-core unemployed. The average annual family income is about the same as that for the urban white families and is well below the OEO Poverty Guidelines. At least 70 percent of the mothers are receiving Welfare, Aid to Dependent Children, or Social Security payments. The average number of children in the family was about the same as in the . . . white families." (20 percent of the Negro mothers had completed high school.)

The medical and nutritional assessments of the 19 target children were extremely extensive and thorough. More than 50 physical signs were checked in the children's medical histories and examinations at the time of the study. Detailed study was made of the children's diets and was compared with the National Nutrition Survey's standards for dietary intakes of calories, proteins, vitamins, and minerals recommended for healthy four to six year old children.

No appreciable nutritional difference was found between the Negro and white samples, and both groups were well <u>above</u> the standards recommended by the National Nutrition Survey. Concerning the results of the medical examinations, the investigators state: "In general, these children were considered to have physical findings within normal limits" (p. 31). In some ways the health conditions of these children were surely not typical of average American children; about half the subjects in each group, for example, had pin worms. But in both groups extremely thorough examination revealed none of the physical or emotional symptoms associated with poor nutrition and usually seen in the studies conducted in Africa, Asia, Mexico, and South

America. With the small samples of this study, the correlations between physical indices and IQ would have too little reliability to be interpretable; they showed no consistent pattern and the authors comment that "We were not surprised that we failed to turn up anything of particular meaning in correlations between the intelligence test scores of the children, and the various indices of skeletal age, height, weight, bone density, and so on. The number of cases was small and only one index of intellectual ability was used. Perhaps with a larger number of cases and increasingly refined techniques of assessment, such relationships might emerge" (p. 61).

Stanford-Binet IQs were obtained on all the children at the conclusion of the study, after they had spent a school year in an intensive experimental nursery school program aimed at improving these children's educability, with particular emphasis on intellectual and motivational factors. The children by this time (5 to 6 years range) were much at ease among teachers and examiners, and were accustomed to interacting with adults in various cognitive games and experimental learning situations. Thus they would seem to be better prepared for Stanford-Binet testing than the normative population in this age range. The mean IQs of the Negro and white groups, respectively, were 76.1 and 95.3 (SDs = 13.0 and 16.7). Here, then, is a considerable IQ difference (more than 1 SD) without there being any appreciable or consistent differences in nutritional status or in physical development and general health.

If signs of malnutrition were not found in these obviously rather extreme socioeconomically disadvantaged groups, the question naturally arises as to what percentage of the United States population, and particularly of the Negro population, suffers from malnutrition to a degree that would affect mental development. Could poor nutrition account for any appreciable fraction of the average white-Negro IQ difference? In order

to gain some outer-bound estimate in answer to this question, I asked Dr. Herbert Birch, a leading researcher in this field, for a rough estimate of the percentage of our population that might suffer a degree of malnutrition sufficient to affect IQ. He said he would guess "Not more than about 1 percent" (personal communication, April 19, 1971). So let us take this figure as the basis for an outer-bound estimate. Assume that all of the 1% of malnutrition in the U.S. population occurs within the Negro population; this would mean that approximately 9% of the Negro population suffers from malnutrition. Assume further that all 9% of this group afflicted by malnutrition has thereby had its IQ lowered by 20 points (which is the difference between malnourished and adequately nourished groups in South Africa -- the most extreme IQ difference reported in the literature). Assuming the present Negro mean IQ in the U.S. to be 85, what then would be the mean if the 20 points of IQ were restored to the hypothetical 9% who had suffered from intellectually stunting malnutrition? It would be 86.70, or a gain of less than 2 IQ points as an outer-bound estimate. Thus it seems unlikely that nutritional factors could carry much weight in any explanation of the average Negro-white IQ difference. This is not to say that cases of malnutrition do not exist in the U.S., or that all possible means should not be applied to ameliorating poor nutrition wherever it was found. It simply means that a nutritional hypothesis of average Negro-white IQ differences has little or no basis in fact, even as a minor contributory factor.

Actually, no one yet knows what the net effect of undernutrition in an entire large population is under natural conditions in which many concomitant factors are free to operate. One might even hypothesize that the net effect of extreme nutritional depression in a population (not for an individual) might actually be beneficial in a eugenic sense, due to increased fetal loss and infant mortality along with natural selection

favoring those who are genetically better endowed physically and mentally. Such a hypothesis could be tesed by analysis of physical and mental measurements on individuals conceived and born during the months of severe protein starvation in various European countries, particularly Holland and Poland, toward the end of World War II. Such studies, sponsored by the U.S. Department of Health, Education, and Welfare, are presently underway.

But there are also less speculative reasons for believing that the role of nutrition should not be overrated as a factor in Negro-white IQ differences. For example, children who are malnourished show a long developmental lag, registered in physical as well as psychological characteristics (Cravioto, 1968). No such lag is found in Negro children and what little evidence there is shows no difference between Negroes and whites in the degree of correlation between physical and mental traits. Malnutrition retards the ossification of cartilage, yet, representative samples of Negro infants have been found to be advanced over whites in ossification (Naylor & Myrianthopoulos, 1967). Malnutrition results in below normal performance on infant tests of sensorimotor development, yet Negro babies generally show advanced performance on these tests as compared with the white norms. Malnutrition impairs memory ability as well as other cognitive functions, yet Negro children show little or no deficit in rote memory. One of the most striking and consistent findings in the research of Cravioto and others is that malnutrition markedly increases the differences between siblings (and conversely lowers sibling correlations) within the same family, both in physical and mental characteristics. I have determined the mean absolute difference between all sibling pairs enrolled in the elementary grades of a California school system on a number of physical and mental measurements, all put on the same scale, with a standard deviation of 15, to make them all comparable to the

IQ scale. The results, shown in Table 5, indicate that there is no appreciable or systematic Negro-white disparity in the magnitudes of the sibling differences and sibling correlations. (The overall Negro-white difference in the value of $|\bar{d}|$ is 0.15 or 0.01 SD.) A nutritional deprivation hypothesis should predict significantly larger sibling differences (and lower correlations) for Negroes than for whites. This prediction clearly is not borne out by the data. Yet these racial groups differ more than 1 SD in both verbal and nonverbal IQ.

Lead Poisoning. This has been hypothesized increasingly of late as a cause of lower Negro IQs. Physical and mental symptoms of lead poisoning typically depend upon the ingestion of excessive quantities of lead over a period of time. Cases of lead poisoning in children are found almost exclusively in those afflicted by pica, a habit of eating non-food substances, occurring most frequently in young children. Nearly all discovered cases of lead poisoning have resulted from children with pica eating the paint peeling off the walls in deteriorating pre-World War II dwellings, usually in urban slums. Almost no post-World War II dwellings have lead paint, and such paint has long been outlawed in the manufacturing of children's toys. Therefore, although the incidence of lead poisoning is not established, it is regarded as a very rare condition as compared with many other health hazards. It has attracted attention largely because of the rather close association that has been found between lead poisoning, pica, and mental retardation. Pica has a much higher incidence among retarded than among normal children, and lead poisoning is highly associated with pica, so the cause-and-effect relationship between lead poisoning and mental retardation remains problematic.

Table 5

Mean Absolute Difference, $|\bar{d}|$, and Correlation, *r*, Between All Siblings of School Age in White and Negro Families on Eleven Measures, Standardized Within 6-month Age Groups, With $\sigma = 15$ for Every Variable in the Combined Populations

Variable	White			Negro		
	N	$|\bar{d}|$	*r*	*N*	$|\bar{d}|$	*r*
Height	1154	12.69	.44	731	12.84	.43
Weight	1155	12.63	.44	731	12.21	.48
Memory I	573	14.52	.27	369	14.25	.29
Memory R	572	13.73	.34	364	15.05	.21
Memory D	568	14.46	.27	358	14.23	.29
Figure Copy	570	13.88	.33	395	13.68	.35
Verbal IQ	1133	13.29	.39	582	12.83	.43
Nonverbal IQ	1132	13.11	.40	600	13.86	.33
Total IQ	341	12.75	.43	200	12.81	.43
Vocabulary	244	12.75	.43	417	12.73	.44
Reading Composition	251	12.93	.42	408	13.89	.33
All Variables		13.34			13.49	

Although the seriousness of lead poisoning, where it occurs, should not be minimized and all possible measures should be taken to prevent its occurrence, the actual known frequencies of the condition appear to be so low that there is no subpopulation whose mean would be snifted one iota by lead poisoning occurring at such low frequencies. Interest in this problem in New York City in recent years has led to increased efforts to discover cases of lead poisoning. In a population of over 8 million, the annual number of reported cases in the mid 1950's was about 100 (Jacobziner, 1966), and this figure rose to 727 in 1969 and 801 in the first half of 1970 (Grince, 1970). Two deaths were attributed to lead poisoning in 1969. The fact that there are many post World War housing areas in which leaded paints have never been used and yet in which the majority of children reared in them have IQs a standard deviation or more below the national average suggests that lead poisoning, though undoubtedly serious when it occurs, is an insignificant factor in relation to average racial or social class differences in IQ and educability.

Reproductive Casualty. The association between social-class, race, and lower IQ, and the much higher incidence of mental retardation among low SES groups, has been attributed in varying degrees to brain impairments incurred prenatally and paranatally. There is a continuum of reproductive casualty, going from fetal and neonatal death to behavioral symptoms referred to as "minimal brain damage." The prevalence of reproductive casualty, most students of the problem agree, is much higher among Negroes than among other groups of similar socioeconomic status. Reproductive casualty is thus frequently mentioned as a major cause of Negro deficits in IQ and scholastic performance. Typical is the statement by Bronfenbrenner (1967, p. 913): "Though the Negro infant is not biologically inferior at the moment of conception, he often becomes so shortly thereafter. The

inadequate nutrition and prenatal care received by millions of Negro mothers result in complications of pregnancy which take their toll in extraordinarily high rates of prematurity and congenital defect. Many of these abnormalities entail neurological damage resulting in impaired intellectual function and behavioral disturbances, including hyperactivity, distractibility, and low attention span. Of particular relevance is the significant role played by paranatal and prenatal factors in the genesis of childhood reading disorders." Statements such as this, it turns out, are extremely difficult, if not impossible, to evaluate on the basis of what is presently known about "reproductive casualty," its causes, its incidence in various subpopulations, and its relationship to mental development. Any reader of the major reviews in this field must be impressed by the chaos and confusion that abounds in this literature and the dearth of consistent and reliable conclusions which can be claimed to have any reasonable degree of generality for any major subpopulation (Amante, et al., 1970; Buck, 1970; Dreger & Miller, 1968, pp. 4-6; Graves, et al., 1970; Hardy, 1965; Pasamanick & Knoblock, 1966; Knoblock & Pasamanick, 1966).

What is quite clear from this literature is that there is some degree of association between prematurity, low birth weight, mother's age (greater risk in early teens and beyond the late thirties), close spacing of pregnancies, and illegitimacy, on the one hand, and higher rates of fetal loss, complications of pregnancy, labor and delivery, infant mortality, neurological difficulties and mental retardation, on the other. It is also clear that both of these sets of conditions have a much higher incidence in the Negro than in the white population. This holds true, in fact, even when Negroes are compared with whites of the lowest SES. The socioeconomically lowest 10 percent of whites in Baltimore, for example, were found to have a 7.6 percent rate of premature births, as compared with 11.4 percent for the full

SES range of the Negro population. The same study reported complications of pregnancy in 14.6 percent of the lower fifth in SES among whites, while for the entire Negro population it was 50.6 percent. The authors state: "These higher rates of prematurity and complications of pregnancy among Negroes over even the lowest white socioeconomic groups are so marked that some workers in this field maintain that they must be attributable to some innate racial characteristic. Since average Negro socioeconomic status is generally lower than that in the lowest white groups, it seems more parsimonious to eliminate the postulated racial factor, and to hypothesize that prematurity and pregnancy complication rates increase exponentially below certain socioeconomic thresholds" (Pasamanick & Knoblock, 1966, p. 19). The authors, however, do not present evidence that the average Negro SES is below the lower tenth or lower fifth of the white population in Baltimore and other studies have found a racial difference in these factors involved in reproductive casualty which are apparently independent of SES (Amante, et al., 1970; Naylor & Myrianthopoulos, 1967).

It is when we begin to evaluate the evidence concerning the relationship of reproductive factors to brain damage and mental development that the real ambiguities arise. Again, there is little doubt that high rates of fetal loss, infant mortality, complications of pregnancy, etc., are epidemiologically associated with higher rates of retardation and brain damage; there are correlations among all these poverty-associated factors. But, surprising as it may seem, what does not emerge clearly from this literature is evidence regarding the direction of causality among these variables. It is not at all clear to what extent the conditions of poverty are themselves a cause of reproductive casualty. Other groups subjected to poverty have not shown high casualty rates on any index. Mechanic (1968) and Graves

et al. (1970) note that Jewish immigrants to America, in spite of their poverty, had even lower rates of infant mortality than any other American group, including the average of the native white population; Orientals are similar in this respect. Something more seems to be involved than just socioeconomic conditions. Amante *et al.* (1970) used a number of signs of CNS dysfunction derived from performance on the Bender Gestalt test to compare Negro and white children in the two lowest SES groups (on a five category scale). The only significant main effect in the analysis of variance was Race ($F = 13.85$, $p < 1$); Social Class and the interaction of Race X Social Class were both nonsignificant ($F < 1$). The authors state: "Rates of brain damage per 100 black children at social class positions IV ($N = 14$) and V ($N = 26$) were, respectively, 50 and 69. . . That is, 50 percent of the class IV children appeared to be brain damaged according to the set of psychological test parameters analyzed, and 69 percent of the class V children. (The corresponding percentages for the white sample were 25 and 26.) Combining classes IV and V into a total sample size of 40 (which of course is still pathetically small), 58 percent of the black children appear to be neurologically handicapped. Further, in the case of the black sample, the frequency of maximal brain damage exceeded the frequency of minimal brain damage -- constituting, therefore, a direct reversal of the minimal-maximal severity pattern observed in the case of the whites. It is apparent, then, that the black population of children is characterized not only by higher overall rates of brain damage relative to the white population, it is also characterized by more severe cases of brain damage" (Amante *et al.*, 1970, p. 126). Later in their discussion of the results, the authors state, "We wish to avoid the implications of racism." And buried among the several following paragraphs discussing the sociology of racism, is put forth what amounts to a major testable hypothesis: "Large

groups of white or black Ss of varying socioeconomic status are typically uncritically selected and tested with conventional psychometric instruments. Naturally, major differences between the classes and the races are indicated. In all probability the observed differences are largely a function of the fact that the groups so selected contain an unspecified number of neurologically handicapped Ss, and these Ss pull the group averages down. If the entire group was neurologically screened to begin with, and the neurologically normal or deviant Ss appropriately compared with other Ss of similar neurological status, and the conventional tests then administered, the supposedly obvious or inevitable IQ differentials between the groups might collapse to zero and/or statistical insignificance" (p. 129).[92] These investigators seem to be ambivalent, however, since they finally state: "Our final conclusion is that there are interclass and interracial differences in terms of measurable intelligence; both environmental and genetic factors appear to contribute to these IQ differentials. At the present time we are assuming that the major factors contributing to group differences are environmental in nature" (p. 129).[93]

At present there are two lines of evidence that seem incompatible with the hypothesis of such a high incidence of brain damage as suggested by Amante et al. and by the many writings of Pasamanick on this subject. The first counterfact is that independently assessed complications of pregnancy are known to be reflected in depressed performance on infant tests of psychomotor development in the first year of life (Honzik, Hutchings, & Burnip, 1965). Yet on these very same tests, given at six months to one year of age, large representative samples of Negro infants were found to do as well as, or better than, comparable samples of white infants (Bayley, 1965a). Such findings could be compatible with a markedly higher incidence of neurological damage in Negro infants only if it is argued that the Negro

infants are normally so very advanced over white infants in psychomotor development that even with a high incidence of brain damage the mean Negro performance is still above the white mean. But this possibility should result in a larger variance of Developmental Quotients for Negroes as compared to whites, and Bayley's data show no significant racial difference in the variance of DQs.

The second item of evidence which is apparently inconsistent with the hypothesis of high rates of brain damage as a principal cause of lower Negro IQ is the heritability of IQ and the intra-family IQ variance (sibling differences) which are about the same for Negro and white populations. If brain damage is an added external source of environmental variance, it should significantly lower the heritability of IQ and increase sibling differences. Negro and white samples which do not differ significantly on these variables still show an IQ difference of 1 SD or more (Scarr, 1971; and see Table 5, in the previous section).

Thus, it is not yet established that the higher rate of reproductive casualty in Negroes, as reflected in a higher incidence of fetal loss, prematurity, and infant mortality, causes any substantial proportion of the IQ deficit. But the question is much too important to be dismissed or allowed to rest ambiguously on the current inadequate state of the evidence. For we do have some good statistics on certain population indices of reproductive casualty, such as fetal loss, and if we make what seems to be a reasonable assumption that fetal death represents merely a threshold effect on a continuous variable of impaired development, we have a firm basis for inferring some higher, though quantitatively undetermined, incidence of physical (including neurological) impairment in the Negro as compared with the white population. In 1965, fetal deaths (for gestation periods of 20 weeks or more) nationwide had almost twice as high a rate among Negroes as

among whites (25.8 vs. 13.3 per 1000 live births) (U.S. Department of Health, Education, and Welfare, 1967, pp. 3-5). Assuming fetal death to be a threshold effect on a normally distributed variable, the Negro and white populations can be said to show a mean difference of 0.46σ on this variable. This is a large difference by any standard. But even if this variable (organismic viability, freedom from impairment, or whatever it is) were perfectly correlated with intelligence, it could account for less than half of the Negro-white IQ difference.

But is the rate of fetal loss in a population entirely a function of external environmental conditions? It appears not to be. The recent research on this matter may provide a clue to the hitherto inexplicably higher rate of fetal loss and other less severe forms of reproductive casualty among Negroes even as compared with non-Negro groups of similar or greater environmental disadvantages.

Bresler (1970) has found that the probability of fetal loss is directly related to the degree of genetic heterogeneity among the ancestral gene pools of the fetus. In two all-white samples from a New England population, comprising all socioeconomic levels, Bresler established highly significant relationships among three factors: fetal loss, the number of countries in the background of parents, and the distances between birthplaces of parents. The ancestry of each fetus, whether lost or live-born, was determined back as far as the great-grandparental generation. Since there are eight great-grandparents, they could have been born in anywhere from 1 to 8 different countries. Bresler determined percentage of fetal loss as a function of the number of different countries among the birthplaces of the great-grand-parental generation. He also determined the percentage of fetal loss as a joint function of number of countries among the great-grandparents and the distance apart (in miles) of the birthplaces of the two parents of the fetus.

Both of these factors serve as indices of the degree of genetic heterogeneity of the fetus's ancestry. Bresler's summary of his main findings: "Data on two white populations show that fetal loss (F_1 generation) in matings of the parental generation (P_1) increases cumulatively by approximately 2.5% to 3% with each additional country of birth in the great-grandparental generation (P_3). A dependent relation shows that increased fetal loss is also related to greater distances between birthplaces of mates within the P_1 generation. Conversely, low fetal loss is encountered with a small number of countries in the background and shorter distance between birthplaces. It is suggested that a large number of countries of birth represents a larger number of Mendelian gene pools and that with increased mixture of these gene pools, fetal loss increases proportionately. An animal model is cited in support of this contention" (p. 24).

Bresler also found that SES had no significant relationship to percentage of fetal loss in these samples. Bresler specifically excluded from his study all families in which one or more persons had any African background in the family history, and he states, "No extrapolation of these findings can be made to interracial matings at this time." So the findings, which are highly reliable, and the genetic theory that explains them can serve only to suggest an hypothesis that the high rate of fetal loss and the various sublethal aspects of reproductive casualty which are a part of this continuum are related to the genetic heterogeneity of the ancestral gene pools of the American Negro.

The genetic interpretation of Bresler's findings is highly technical and cannot be elaborated upon here, but it has been tested and reliably demonstrated in numerous animal experiments (referenced by Bresler, 1970). Briefly, more distantly related gene pools have greater genetic imbalance between gene loci on the chromosomes; the loci for certain genes do not

match up properly, so that if the two alleles required for the production of an enzyme have undergone some evolutionary translocations on the chromosomes, the enzyme controlled by a particular gene may not be produced and therefore cannot make its necessary contribution to the normal development of the growing embryo or fetus. Different genes become important at various stages of development, and some genetic imbalances will prove lethal while others will be sublethal but can cause developmental anomalies of varying severity. The effects have been demonstrated, for example, with frogs, all of the same species, but distributed over a wide geographical range. Bresler (1970, p. 24) summarizes some of the findings from these experiments, in which genetic crosses are made between frogs of the same species collected from varying geographic distances:

1. The hybrids between members of adjacent geographical territories tended to be normal in development and morphology.

2. The greater the geographical distance between parental combinations in eastern North America, the more retarded was the rate of development, the greater were the morphological defects in the hybrids, and the fewer were the normal individuals.

3. The greater the geographical distance between parental combinations, the larger was the percentage of eggs which failed to develop properly.

4. The further apart in geographical distance . . . the members were collected from, the earlier in development did reproductive wastage occur.

What about heterosis,or hybrid vigor, which usually results from outcrossings? Heterosis results when there are dominant and recessive genes involved in a characteristic. Outcrossing increases the likelihood of heterozygotes, and if the dominant genes are "desirable," hybrid vigor (e.g., greater size) is said to result. The effect is quite independent

of the effects of loci imbalance involved in the Bresler study; it is a different phenomenon entirely. The two effects can operate simultaneously, in "opposite" directions; the deleterious effects of genetic imbalance can override desirable effects of heterosis. The opposite of heterosis genetically is inbreeding depression, which shows up most clearly in consanguineous matings, such as cousin marriage. Not all traits show heterosis and inbreeding depression. Height and chest circumference show it, as does IQ (Schull & Neel, 1965), but head shape (cephalic index) and circumference apparently do not (Wolański, Jarosz, & Pyżuk, 1970). Too close inbreeding causes depression of some characteristics because of the increased likelihood of the pairing of undesirable mutant alleles, while too much heterogeneity of ancestral gene pools can have undesirable consequences due to genetic imbalance caused by translocations and inversions of loci.

The role of these genetic mechanisms in the causation of reproductive casualty and its differing rates in various subpopulations calls for much further investigation, which hopefully will not be hindered by ideologically motivated dogmatic insistence that all such effects must be attributable entirely to external environmental factors.

FOOTNOTES

* adopting parents. In the frequently cited study by Skodak and Skeels (1949), children of rather low IQ mothers (mean = 85.75) were adopted into superior foster homes. They showed a correlation of .38 with their true mothers with whom they had no contact beyond infancy. The adopted children's average IQ, however, was approximately 11 points higher than the mean IQ that would be predicted from a genetic model assuming that the children represented a random selection of the offspring of mothers with a mean IQ of 85 and were placed in randomly selected environments in the population. Actually, of course, these children were selected by the adoption agency as suitable for adoption and the adoptive homes were selected for their favorable environmental attributes. The 11 points, however, is very likely an overestimate of any environmental effect on these children's IQs, since the children put out for adoption, most of them illegitimate, were not a random selection of such children, and it has been indicated by Leahy (1935) that illegitimate children who become adopted have a higher average IQ than illegitimate children in general or than legitimate children placed for adoption. Readers interested in a detailed and trenchant critique of the Skodak and Skeels studies should read Terman (1940, pp. 462-467) and McNemar (1940).

[70]The Negro and white populations of the U. S. today differ about 1 SD in SES in terms of the SD of SES in the white population. Thus the average SES difference between the races is approximately the same as the average absolute difference among persons within the white population.

[71]The mean IQs of the white upper and lower SES groups in Tulkin's study are at the 81st and 34th percentiles of the white population norms; the corresponding percentiles of the upper and lower SES Negro samples, based on Negro population parameters ($\bar{X}$ = 85, σ = 14) are 95th percentile and 66th percentile, respectively.

*Ed. Note: Footnote 69 copy was not received in time for inclusion in printing; for ref. see author.

[72]Shuey (1966, p. 520, footnote 55) gives the following means (of children's scores) for the combined studies:

Socioeconomic Status

Children	Upper	Lower	Difference
White	111.88	94.22	17.66
Negro	91.63	82.04	8.59
Difference	20.25	12.18	

Assuming a single parent-offspring regression of 0.50 and no assortative mating, which is the simplest possible genetic model, and assuming a white population mean of 100 and a Negro population mean of 85, the mean IQs of the most extreme parent (probably the father or the one who chiefly determines the family's SES) are estimated as follows:

Socioeconomic Status

Parent	Upper	Lower	Difference
White	124	88	36
Negro	99	79	20
Difference	25	9	

It can be seen that the race x SES crossover in children's IQs (shown by the diagonal in the top table) must result because the upper SES white and Negro parents differ quite markedly in IQ, assuming a genetic interpretation of Shuey's data is correct. If the Negro parent mean IQs were perfectly matched to the white IQs, this simple genetic model, given the assumptions previously stated, would predict mean IQs of 104.5 and 86.5 for the upper and lower SES Negro children.

It also appears that, in terms of IQ, the high SES white samples in those studies summarized by Shuey may represent a slightly more select upper segment of the white population while the high SES Negro samples may represent a somewhat less select upper segment of the Negro population. The white children's mean IQ of 111.88 is at approximately the 73rd percentile in the white population while the Negro children's mean of 91.63 is only at about the 70th percentile in the Negro population, assuming equal σ in both populations. If the Negro σ is smaller than the white σ, as is true in the majority of studies, then the high SES Negro samples in Shuey's summary could be a more select segment of the Negro population than is true of the high SES white sample.

[73]The environmental variables were: (1) reading material in home, (2) items in home (cultural amenities), (3) structural integrity of home, (4) foreign language in home, (5) preschool attendance, (6) encyclopedia in home, (7) parents' education, (8) time spent on homework, (9) parents' educational desires for child, (10) parents' interest in school work, (11) child's self concept (self-esteem), (12) child's interest in school and reading.

[74]The largest and methodologically most thorough study of this question showed that racial composition of the classroom of itself had no effect on IQ (Wilson, 1967).

[75]If from a normal distribution, with mean = 0, σ = 1, a segment of the distribution lying between two points on the abscissa, $\underline{z}_1$ and $\underline{z}_2$, is eliminated, the resulting mean ($\bar{X}_s$) of the eliminated segment is:

$$\bar{X}_s = \frac{\underline{Y}_1 - \underline{Y}_2}{\text{area between } \underline{z}_1 \text{ and } \underline{z}_2}$$

where $\underline{Y}_1$ and $\underline{Y}_2$ are the values of the ordinate at $\underline{z}_1$ and $\underline{z}_2$. The mean of one tail of a distribution then is simply

$$\bar{X}_t = \frac{Y_1}{\text{area beyond } \underline{z}_1}$$

[76]There is one peculiarity to be noted in this study and about which the authors make no comment. The mean IQs of all groups are unusually high (overall mean = 112.15). While the IQ difference between low and middle SES groups was by far the largest effect in the experiment, the SES IQ difference was still smaller than is generally found. The mean posttest IQs of low and middle SES groups were 106.86 and 117,18, respectively. Nothing is said about the racial composition of the samples, but a mean IQ of 106.86 is certainly well above that generally found for either white or Negro samples classed as "low SES." This atypical peculiarity of the data in an otherwise impeccable study may limit its generalizability to more typical populations.

[77]Sattler's summary: "Katz and his co-workers' series of studies show that white testers in comparison with Negro testers do not necessarily impede the performance of Negro college subjects. Negro testers obtained significantly higher scores than white testers when the probability-of-success conditions informed the college subjects that they had little or an equal chance of equaling the white norm; white testers obtained significantly higher scores than Negro testers under certain conditions and with certain groups (e.g., motor instructions and hard digit-symbol task; digit-letter substitution and mild threat; college students with satisfactory high school averages in any probability of success condition)" (p. 143).

[78]A convenient index for expressing the verbal-nonverbal discrepancy, suggested by Weyl (1969), is 100 x nonverbal/verbal score (expressed in

standard score form). Applied to the Coleman data at 1st grade, before the schools could have had any appreciable cumulative effect, this index for the various subpopulations is Puerto Rican 102.0, Indian 110.9, Mexican 107.7, Oriental 109.7, White 101.6, Negro 95.6.

[79]Apparently the sensorimotor capability for manipulating objects in the environment also is not crucial for normal cognitive development. The children who are probably the most disadvantaged in this respect are the limbless thalidomide babies, who, despite their severe motor handicap, are reported to show no deficit in cognitive development (Bower, 1971).

[80]I am indebted to Dr. Neil Warren, University of Sussex, for several additions to this list of references, particularly the unpublished studies by Kilbride (1969), Theunissen (1948), and Falmagne (1959). Dr. Warren has prepared a valuable review of the literature on this topic ("African Infant Precocity," unpublished manuscript, 1971). His conclusions are summarized in his own abstract of this review: "Studies of African infant development are reviewed, with reference to the possible phenomenon of African infant precocity, and in an attempt to place these studies within the perspective of a viable strategy of cross-cultural research. Although the majority of studies report precocity, it is held that defects of measurement and design must preclude the conclusion that precocity is an established fact. The better-designed studies do not report precocity. It is also argued that infant differences by social milieu afford the most sensible basis for the necessary introduction of independent variables into this research area, and that improved techniques of assessment should be applied, both in the neonatal period and beyond." The "better-designed" studies referred to are the two unpublished studies (Theunissen, 1948; Falmagne, 1959); they report no evidence of infant motor precocity in their African

samples (which were from different parts of African than those of other studies). I have not had the opportunity to examine these unpublished theses at first hand, but it is interesting that, as the only studies which have not reported advanced development in Negro infants, Warren seems to regard them as the only studies which are methodologically sound. I do not concur in this conclusion. Though many of the other studies surely cannot be held up as methodological paragons of rigorous measurement and statistical inference, the striking magnitude of the differences observed and the great consistency of the many studies by different investigators using various techniques lends a weight to the preponderance of evidence which cannot be dismissed by two studies, whatever their methodological excellence, based on different African subpopulations. In such a case, the difference in results is much more likely due either to sampling differences or to true subpopulation differences, rather than to methodological faults which would have caused all other studies to yield opposite conclusions. One almost wonders if Dr. Warren's rather extreme weighting of the evidence toward the weakest possible conclusion is an illustration of Bertrand Russell's remark that "an intransigent perfectionism is the last refuge of the skeptic." It seems apparent that the overall consistency and convergence of many lines of evidence which point in the same direction must have been largely ignored by Dr. Warren in summarizing his conclusions. He does not mention studies (e.g., Naylor & Myrianthopoulos, 1967; Nelson & Dean, 1959) which show Negro infants' advanced development in physical characteristics such as rate of bone development (determined from X-rays showing the rate of ossification of cartilege) and the greater maturity of brain wave patterns seen in electroencephalograms. Nor is there mention of those American studies, with the exception of Bayley's (1965), which are methodologically sound and more sophisticated than most of the African studies (Durham Education Improvement

Program, 1966-67a,b; Knoblock & Pasamanik, 1953; Pasamanik, 1946; Walters, 1967; Williams & Scott, 1953), and all of which report advanced motor development of Negro as compared with white infants. Of all existing studies, including the unpublished studies referred to by Dr. Warren, Bayley's (1965) is based on the largest (1,409 infants) samples of Negroes and whites (and also the most representative of the U.S. Negro and white populations). The standardization and carefulness of testing procedures and the complete adequacy of the presentation of data and the statistical analyses thereof make it probably the best single study available to date. Summarizing the results on the Developmental Motor Quotient (DMQ), derived from her infant tests, Bayley (1965, p. 405) reports: "The means for the Negroes are higher at every age (from 1 to 15 months) except 15 months. The difference reaches significance at the .01 level of confidence at months 3, 4, 5, and 9 and at the .05 level at months 7 and 12. After 12 months this difference disappears." Bayley concludes: "Although there is considerable overlap of scores among whites and Negroes of the same age, a genetic factor may be operating. That is, Negroes may be inherently more precocious than whites in their motor coordinations" (pp. 408-9). Cravioto (1966, p. 78) has noted similar results in Indian infants of Guatemala and Mexico, commenting that on the Gesell tests of infant behavior "their development at two or three weeks is similar to that of Western European infants two or three times as old." It is also interesting that Orientals (Chinese Americans) who, as school age children, equal or exceed the white population in the most heavily *g* loaded intelligence tests and in the most abstract scholastic subjects, as infants are significantly less motorically reactive than white infants, though they show no significant difference in neuromuscular maturity *per se* (Freedman & Freedman, 1969). Chinese and Caucasian neonates in the nursery of the same hospital were "tested" shortly after birth. Marked differences in reactivity

showed up, for example, when a loosely woven cloth was placed on the face of the supine baby. The typical Caucasian neonate "immediately struggled to remove the cloth by swiping his hands and turning his face;" the typical Chinese-American neonate "lay impassively, exhibiting few overt motor responses." This behavioral difference was significant beyond the .0001 level! "Similarly, when placed in a prone position, the Chinese infants frequently lay as placed, with face flat against the bedding, whereas the Caucasian infants either turned the face to one side or lifted the head." Similar studies of Negro neonates suggest that the three racial groups lie on a developmental continuum on which the Caucasian group is more or less intermediate.

[81]The basic standard score, called a *z* score, for an individual is simply

$$\underline{z} = (X - \bar{X}) \,/\, \underline{SD}$$

where X is the individual's raw score on the test, $\bar{X}$ is the mean raw score of the standardization sample (or some precisely defined age group within the sample), and *SD* is the standard deviation of the sample raw scores. This *z* score can then be transformed to any convenient scale, with a mean of *M* and a standard deviation of σ.

Transformed $\underline{z}' = \sigma \underline{z} + \underline{M}$. A so-called *T* scale has $\underline{M} = 100$, $\sigma = 10$. On the conventional IQ scale, $\underline{M} = 100$, $\sigma = 15$.

[82]For an excellent, though quite technical, discussion of the methodological aspects of this issue, the reader is referred to Einhorn and Bass (1971).

[83]I am indebted to Dr. Mabel C. Purl, Director of Research and Evaluation, Riverside Unified Schools, for these data.

[84]I did a principal components analysis of the correlation matrix in Tuddenham's Table 3.1 (p. 66), which in addition to the ten Piagetian tests contains the variables Age, Sex, and Father's Occupation. On the first principal component (the general factor common to all the tests), father's occupation has a loading of 0.46. The average loading of the 10 Piagetian items is .51, with a range from .22 (lateral reversal) to .76 (conservation of volume).

[85]If we assume the conservation tests reflect an underlying normal distribution of cognitive development, these percentages correspond to a white-Negro difference of between 1 and 1.5σ, which is the range of difference generally found between low SES Negro and middle SES white children on conventional IQ tests.

[86]Highly precise tachistoscopic equipment is required for this work. For technical details of the phenomenon and its measurement the reader is referred to Holland (1963).

[87]For information theory, the "bit" (an abbreviation for <u>binary digit</u>) is the basic unit for quantifying information. One "bit" is the amount of information necessary to resolve two equally probable alternatives; it is equivalent to the minimum number of binary questions (answerable with Yes or No) needed to reduce the uncertainty to zero. For example, a "problem" involving no more than making one choice from among one alternative contains zero information; if there are two alternatives, there is 1 bit of information, since one must receive the answer to only one binary question in order to know which choice is "correct"; if there are 4 alternatives, there are 2 bits of information; 8 alternatives = 3 bits. The number of bits ($\underline{n}$) can be seen in the following relationship $\sqrt[n]{\text{alternatives}} = 2$.

[88]The locations of these studies: Cape Town, South Africa; Kampala, Uganda; Guatemala; Mexico; Santiago, Chile; Sarajevo, Yugoslavia; Hyderabad, India; Peru.

[89]In autopsy studies of stillborn and newborn infants of poor, presumably undernourished mothers in New York City, as compared with infants of non-poor mothers, the magnitude of the effects of "poorness" (presumably maternal undernutrition) on the growth of various organs and body measurements was determined. Of the 8 measurements made on the babies, the brain was least affected, suggesting that it is probably the nutritionally most highly buffered organ in the fetus (Naeye, Diener, Dellinger, & Blanc, 1969). The index of relative effect of prenatal undernutrition for the 8 infant body measurements and the placenta were: thymus 38, adrenals 25, spleen 23, heart 15, body length 15, liver 21, kidney 10, brain 6, and placenta 4.

[90]The closest to anything like this that I have been able to find in the relevant literature on human nutrition and growth is a study carried out in Aberdeen, Scotland which showed a correlation between the mother's nutritional status as determined by her height and her infant's birth weight and subsequent growth and development; there were also correlations between mother's height, child's height at age 7, and his performance on achievement tests in school (Birch, Richardson, Baird, Horobin, & Illsley, 1970). The possible confounding of genetic and nutritional factors in determining these relationships rules out any definite interpretation.

[91]The investigation also conducted parallel studies in two rural samples of poor Appalachian whites in Tennessee. Their nutritional status was higher than that of the urban samples mainly because they raised much of their own food. About 70 percent of the rural samples, as compared with

about 50 percent of the urban samples, had pin worms. The IQs of the rural whites averaged about 6 points lower than of the urban whites.

[92]If we assume a mean IQ of 100 in the white populations, 85 in the Negro population, and a SD of 15 in each population; and further assume that all brain damaged persons are of lower IQ than neurologically non-damaged persons, what percentage of the Negro population would have to be screened out as neurologically damaged in order to bear out this prediction by Amante, et al.? Calculations indicate that at least the lower 62 percent of the Negro population in IQ would have to be screened out for the remaining 38 percent to have a mean IQ of 100. This surprising conclusion would seem to cast some doubt on this hypothesis. Have there been any estimates of the incidence of neurological damage in the Negro population which come anywhere near such a figure and which are also experimentally independent of the IQ determination, i.e., based on criteria that does not include measurement of the dependent variable, viz., IQ? I know of none. The criteria used in the study by Amante et al. were derived from signs of the Bender Gestalt Test, which itself has a high g loading, making it difficult, if not impossible, to distinguish reliably between signs of neurological impairment and sheer mental immaturity. Among neurologically normal white middle-class elementary school children, Bender-Gestalt performance would correlate substantially with conventional measures of IQ. The white-Negro difference found by Amante, et al. on the Bender Gestalt, however, would be roughly equivalent to 11 IQ points, which is slightly larger than the average difference typically found for Negro and white children within the lower social strata.

[93]It is interesting to see this conclusion of genetic interclass and interracial IQ differences appearing in a 1970 issue of Journal of Social Issues, the official periodical of the Society for the Psychological Study

of Social Issues, the Council of which in 1969 publicly censured me for drawing a similar conclusion in my article in the Harvard Educational Review. My statement was: "The preponderance of the evidence is, in my opinion, less consistent with a strictly environmental hypothesis than with a genetic hypothesis, which, of course, does not exclude the influence of environment or its interaction with genetic factors" (Jensen, 1969*a*, p. 82).

References

Ainsworth, M. D. S. The development of infant-mother interaction among the Ganda. In Foss, B. M. (Ed.), Determinants of infant behavior. Vol. 2. London: Methuen, 1963.

Ainsworth, M. D. S. Infancy in Uganda. Baltimore: John Hopkins University Press, 1967.

Amante, D., Margules, P. H., Hartman, Donna M., Storey, Dolores B., & Lewis, J. W. The epidemiological distribution of CNS dysfunction. Journal of Social Issues, 1970, 26- 105-136.

Ames, Louise B., & Ilg, Frances L. Search for children showing academic promise in a predominantly Negro school. Journal of Genetic Psychology, 1967, 110, 217-231.

Anastasi, Anne. Heredity, environment, and the question "How?" Psychological Review, 1958, 65, 197-208.

Anderson, J. G., & Johnson, W. H. Stability and change among three generations of Mexican-Americans: Factors affecting achievement. American Educational Research Journal, 1971, 8, 285-309.

Andor, L. E. Aptitudes and abilities of the black man in sub-Saharan Africa 1784-1963: An annotated bibliography. Johannesburg: National Institute for Personnel Research, 1966.

Baratz, Joan C. Teaching reading in an urban Negro school system. In Williams, F. (Ed.) Language and Poverty: Perspectives on a theme. Chicago: Markhan, 1970. Pp. 11-24.

Bayley, Nancy. Comparisons of mental and motor test scores for ages 1-15 months by sex, birth order, race, geographical location, and education of parents. Child Development, 1965, 36, 379-411. (a)

Bayley, Nancy. Learning in adulthood: The role of intelligence. In H. J. Klausmeier & C. W. Harris (Eds.) Analyses of concept learning. New York: Academic Press, 1966. Pp. 117-138.

Bereiter, C. Genetics and educability: Educational implications of the Jensen debate. In Hellmuth, J. (Ed.) Disadvantaged child. Vol. 3 Compensatory education: A national debate. New York: Brunner-Mazel, 1970. Pp. 279-299.

Berry, J. W. Temne and Eskimo perceptual skills. International Journal of Psychology, 1966, 1, 207-222.

Birch, H. G. Problems inherent in population studies of nutrition and mental subnormality. In Jervis, G. A. (Ed.), Expanding concepts in mental retardation. Springfield, Ill.: Charles C. Thomas, 1968.

77-843 O - 72 - 27

Birch, H. G., Richardson, S. A., Baird, S. D., Horobin, G., & Illsley, R. Mental subnormality in the community: A clinical and epidemiologic study. Baltimore: Williams & Wilkins, Co., 1970.

Blalock, H. M.,Jr. Social Statistics. New York: McGraw-Hill, 1960.

Bloom, B. S. Stability and change in human characteristics. New York: Wiley, 1964.

Bloom, B. S. Letter to the editor, Harvard Educational Review, 1969, 39, 419-421.

Bodmer, W. F., & Cavalli-Sforza, L. L. Intelligence and race. Scientific American, 1970, 223, No. 4, 19-29.

Bosco, J. J. Social class and the processing of visual information. Final Report, Project No. 9-E-041, Contract No. OEG-5-9-325041-0034(010). Office of Education, U.S. Dept. of Health, Education, and Welfare, May, 1970.

Bower, T. G. R. Early learning and behavior. The biology of growth--1. Times Literary Supplement, May 7, 1971, pp. 523-524.

Boyer, W. H., & Walsh, P. Are children born unequal? Saturday Review, October 19, 1968, pp. 61-63, 77-79.

Bracht, G. H. Experimental factors related to aptitude treatment interactions. Review of Educational Research, 1970, 40, 627-645.

Bresler, J. Outcrossings in Caucasians and fetal loss. Social Biology. 1970, 17, 17-25.

Bronfenbrenner, U. The psychological costs of quality and equality in education. Child Development, 1967, 38, 909-925.

Brown, R. Social psychology, New York. Free Press, 1965.

Buck, Carol. Examples of current studies of reproductive casualty. In Haywood, H. C. (Ed.), Social-cultural aspects of mental retardation. New York: Appleton-Century-Crofts, 1970.

Burgess, Johanna, & Jahoda, Marie. The interpretation of certain data in "How Much Can We Boost IQ Score and Scholastic Achievement," Bulletin of the British Psychological Society, 1970, 23, 224-225.

Burks, B. S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent--foster child resemblance and true parent--true child resemblance. Yearbook of the National Society for the Study of Education, 1928, 27 (I), 219-316.

Burt, C. Ability and income, British Journal of Eudcational Psychology, 1943, 13, 83-98.

Burt, C. Class differences in general intelligence: III. British Journal of Statistical Psychology, 1959, 12, 15-33.

Burt, C. The gifted child. British Journal of Statistical Psychology, 1961, 14, 123-139. (a)

Burt, C. Intelligence and social mobility. British Journal of Statistical Psychology, 1961, 14, 3-24.(b)

Burt, C. Is intelligence distributed normally? British Journal of Statistical Psychology, 1963, 16, 175-190.

Burt, C. The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. British Journal of Psychology, 1966, 57, 137-153.

Burt, C., & Howard, M. The multifactorial theory of inheritance and its application to intelligence. British Journal of Statistical Psychology, 1956, 9, 95-131.

Burt, C., & Williams, E. L. The influence of motivation on the results of intelligence tests. British Journal of Statistical Psychology, 1962, 15, 127-136.

Butcher, H. J. Human intelligence: Its nature and assessment. London: Methuen, 1968.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Carter, J., Gilmer, Barabara, Vanderzwaag, R., & Massey, Katherine. "Health and nutrition in disadvantaged children and their relationship with intellectual development." Collaborative Research Report. Vanderbilt School of Medicine, Nashville, Tennessee, 1971.

Cattell, R. B. Personality. New York: McGraw-Hill, 1950.

Clark, K. The cult of cultural deprivation: A complex social phenomenon. In Natchez, G. (Ed.), Children with reading problems, 1968.

Coleman, J. S., et. al. Equality of educational opportunity. U. S. Office of Education, Washington, D. C., 1966.

Cooper, R., & Zubeck, J. Effects of enriched and restricted early environments on the learning ability of bright and dull rats. Canadian Journal of Psychology, 1958, 12, 159-164.

Council of the Society for the Psychological Study of Social Issues. Statement by SPSSI on current IQ controversy: Heredity versus environment. American Psychologist, 1969, 24, 1039-1040.

Cravioto, J. Malnutrition and behavioral development in the preschool child. Pre-School child malnutrition, National Health Science, Public. No. 1282, 1966.

Cravioto, J. Nutritional deficiencies and mental performance in childhood. In Glass, D. C. (Ed.), Environmental influences. New York: Rockefeller University Press, 1968.

Cravioto, J., & Delicardie, Elsa R. Mental performance in school age children: Findings after recovery from early severe malnutrition. American Journal of Diseases of Children, 1970, 120, 404-410.

Cronbach, L. J. Heredity, environment, and educational policy. Harvard Educational Review, 1969, 39, 338-347.

Cronbach, L. J., & Snow, R. E. Final report: Individual differences in learning ability as a function of instructional variables. Stanford, University, March 1969.

Crow, J. F. Genetic theories and influences: Comments on the value of diversity. Harvard Educational Review, 1969, 39, 301-309.

Crow, J. F. Do genetic factors contribute to poverty? In Allen, V. L. (Ed.) Psycholgocial Factor in poverty. Chicago: Markham, 1970. Pp. 147-160.

Crow, J. F., & Felsenstein, J. The effect of assortative mating on the genetic composition of a population. Eugenics Quarterly, 1968, 15, 85-97.

Curti, M., Marshall, F. B., Steggerda, M., & Henderson, E. M. The Gesell schedules applied to one-, two-, and three-year old Negro children of Jamaica, B. W. J. Journal of Comparative Physiological Psychology, 1935, 20, 125-156.

DeLemos, Marion M. The development of conservation in aboriginal children. International Journal of Psychology, 1969, 4, 255-269.

Deutsch, Cynthia, & Deutsch, M. Theory of early childhood environment programs. In R. Hess and B. Bear (Eds.), Early education: current theory, research action. Chicago: Aldine Publishing Co., 1968.

Deutsch, M. Happenings on the way back to the forum: Social science, IQ, and race differences revisited. Harvard Educational Review, 1969, 39, 523-557.

Dobbing, J. Undernutrition and the developing brain: The relevance of animal models to the human problems. American Journal of Diseases of Children, 1970, 120, 404-410.

Dreeben, R. Comments on Jensen. Administrator's Notebook, Midwest Administration Center, University of Chicago, 1969, 18, no. 3.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. Psychological Bulletin, 1960, 57, 361-402.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and white in the United States: 1959-1965. Psychological Bulletin Monograph Supplement, 1968, 70, No. 3, Part 2.

Duncan, O. D. Ability and achievement. Eugenics Quarterly, 1968, 15, 1-11.

Duncan, O. D. Inequality and opportunity. Presidential address at the annual meeting of the Population Association of America, Atlantic City, N. J., April 11, 1969.

Durham Education Improvement Program, 1966-1967. (a)

Durham Education Improvement Program, Research, 1966-1967. (b)

Eckland, B. K. Genetics and sociology: A reconsideration. American Sociological Review., 1967, 32, 173-194.

Eichenwald, H. F., & Fry, Peggy C. Nutrition and learning. Science, 1969, 163, 644-648.

Einhorn, H. J., & Bass, A. R. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.

Eisenberg, L., Berlin, C., Dill, Anne, and Sheldon, F. Class and race effects on the intelligibility of monosyllables. Child Development, 1968, 39, 1077-1089.

Entwisle, Doris R. Semantic systems of children: Some assessments of social class and ethnic differences. In Williams, F. (Ed.), Language and poverty: Perspectives on a theme. Chicago, Markham, 1970. Pp. 123-139.

Evans, Judith L. Children in Africa: A review of psychological research. New York: Teachers College Press, Columbia University, 1970.

Faladé, S. Contribution a une etude sur le developpement de l'enfant d' Afrique noire. Paris: Foulon, 1955.

Faladé, S. Le developpement psycho-moteur de l'enfant Africain due Senegal. Concours medical, 1960, 82, 1005-1013.

Falconer, D. S. An introducation to quantitative genetics. New York: Ronald Press, 1960.

Falmagne, J. Ce. Étude de certain aspects du developpement du nourrisson de 0 à 6 mois. Unpublished thesis, Université Libre de Bruxelles, Bruxelles, 1959.

Flavell, J. The developmental psychology of Jean Piaget. New York: Van Nostrand, 1963.

Fleming, Elyse S., & Auttonen, R. G. Teacher expectancy or my fair lady. American Educational Research Journal, 1971, 8, 241-252.

Freedman, D. G., & Freedman, Nina Chinn. Behavioral differences between Chinese-American and European-American newborns. Nature, 1969, 224, 1227.

Freeman, H. E., Ross, J. M. Armor, D., & Pettigrew, T. F. Color graduation and attitudes among middle income Negroes. American Sociological Review, 1966, 31, 365-374.

Furth, H. G. Research with the deaf: Implications for language and cognition. Psychological Bulletin, 1964, 62, 145-164.

Galton, F. Hereditary genius (2nd ed.) New York: Appleton, 1870.

Gaudia, G. Race, social class, and age of achievement of conuervation on Piaget's tasks. Developmental Psychology, in press.

Géber, M. Développement psycho-moteur de l'enfant Africain. Courrier, 1956, 6, 17-29.

Géber, Marcelle. The psycho-motor development of African children in the first year, and the influence of maternal behavior. Journal of Social Psychology, 1958, 47, 185-195. (a)

Géber, M. L'enfant africain occident alisé et de niveau social superieur en Ouganda. Courrier. 1958, 8, 517-523. (b)

Géber, M. Problems posés par le développement du jeune infant africain en foution de so milieu social. Travail Humain, 1960, 1-2, 97-111.

Géber, M. Longitudinal study and psychomotor development among Baganda children. In Proceedings, Fourteenth International Congress of Applied Psychology, Vol. 3. Munsgaard: Copenhagen, 1962.

Géber, Marcelle, & Dean, R. F. A. The state of development of newborn African children. Lancet, 1957, 1216-1219. (a)

Géber, M., and Dean, R. F. A. Gesell tests on African children. Pediatrics, 1957, 20, 1055-1065. (b)

Géber, M., & Dean, R. F. A. Psychomotor development in African children: the effects of social class and the need for improved tests. Bulletin of the World Health Organization, 1958, 18, 471-476.

Géber, M., & Dean, R. F. A. Le développement psycho-moteur et somatique des jeunes enfants africains en Ouganda. Courrier, 1964, 15, 425-437.

Géber, M., & Dean, R. F. A. Precocious development in newborn African infants. In Brackbill & Thompson (Eds.)Readings in infancy and childhood, New York: Free Press, 1966.

Ghiselli, E. E. Managerial talent. American Psychologist, 1963, 18, 631-642.

Gilliland, A. R. Socioeconomic status and race as factors in infant intelligence test scores. Child Development, 1951, 22, 271-273.

Gittell, Marilyn. New York City School Fact Book. Institute for Community Studies, Queens College, Flushing, Long Island, New York, 1971.

Glass, B., & Li, C. E. The dynamics of racial intermixture--An analysis based on the American Negro. American Journal of Human Genetics, 1953, 5, 1-20.

Glassman, Lynette D. A social-class comparison of conceptual processes in children's free recall. Unpublished doctoral dissertation, University of California, 1968.

Goldhammer, H. Letters. Science, 1971, 172, 10.

Gordon, E. W. Problems in the determination of educability in populations with differential characteristics. In Hellmuth, J. (Ed.) Disadvantaged Child, Vol. 3. Compensatory education: A national debate. New York: Brunner/Mazel, 1970. pp. 249-267.

Gottesman, I. I. Biogenetics of race and class. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.) Social class, race, and psychological development. New York: Holt, Rinehart, & Winston, 1968. Pp. 11-51.

Gough, H. G. A short social status inventory. Journal of Educational Psychology, 1949, 40, 52-56.

Gough, H. G. A nonintellectual intelligence test. Journal of Consulting Psychology, 1953, 17, 242-246.

Gough, H. G. A cluster analysis of Home Index status items. Psychological Reports, 1971, 28, 923-929.

Graves, W. L., Freeman, M. G., & Thompson, J. D. Culturally-related reproductive factors in mental retardation. In Haywood, H. C. (Ed.), Social-Cultural aspects of mental retardation. New York: Appleton-Century-Crofts, 1970.

Gross, M. Learning readiness in two Jewish Groups. New York: Center for Urban Education, 1967.

Guince, V. F. Statement on lead poisoning. Presented to Sub-committee on Housing of the House Committee on Banking and Currency, July 22, 1970.

Haldane, J. B. S. The implications of genetics for human society. In Geertz, S. J. (Ed.) Genetics Today: Proceedings of the XI International Congress of Genetics. The Hague, Netherlands, September, 1963. New York: Pergamon Press, 1965, pp. xcii-xciii.

Hardy, Janet B. Perinatal factors and intelligence. In Sonia F. Osler & R. E. Cooke (Eds.) The biosocial basis of mental retardation. Baltimore, Md.: The Johns Hopkins Press, 1965, Pp. 35-60.

Harms, L. S. Listener comprehension of speakers of three status groups. Language and Speech, 1961, 4, 109-112.

Harrell, R. F., Woodyard, E., & Gates, A. I. The effects of mothers' diets on the intelligence of offspring. New York: Bureau of Publications, Teachers College, 1955.

Harris, A. J., & Lovinger, R. J. Longitudinal measures of the intelligence of disadvantaged Negro adolescents. School Review, 1968, 76, 60-66.

Harrison, G. A. The measurement and inheritance of skin colour in men. Eugenics Review, 1957, 49, 73-76.

Harrison, G. A., & Owen, J. J. T. The application of spectrophotometry to the study of skin color inheritance. Acta Genetica, 1956, 4, 480-485.

Harrison, G. A., & Owen, J. J. T. Studies on the inheritance of human skin color. Annals of Human Genetics. 1964, 28, 27-37.

Harrison, G. A., Owen, J. J. T., Da Rocha, F. J., & Salzano, F. M. Skin color in Southern Brazilian populations. Human Biology, 1967, 39, 21-31.

Haseman, J. K., & Elston, R. C. The estimation of genetic variance from twin data. Behavior Genetics, 1970, 1, 11-19.

Havighurst, R. J. Minority subcultures and the law of effect. American Psychologist, 1970, 25, 313-322.

Herskovits, M. J. On the relation between Negro-white mixture and standing in intelligence tests. Pediatric Seminar, 1926, 33, 30-42

Heston, L. L. Psychiatric disorders in foster home reared children of schizophrenic mothers. British Journal of Psychiatry, 1966, 112, 819-825.

Hirsch, J. Behavior-genetic analysis and its biosocial consequences. Seminars in Psychiatry, 1970, 2, 89-105.

Holland, H. C. "Visual masking" and the effect of stimulant and depressant drugs. In Eysenck, H. J., (Ed.) Experiments with drugs. New York: Pergamon, 1963.

Holzinger, K. J. The relative effect of nature and nurture influences on twin differences. Journal of Educational Psychology, 1929, 20, 241-248.

Honzik, M. P. Developmental studies of parent-child resemblance in intelligence. Child Development, 1957, 28, 215-228.

Honzik, Marjorie P., Hutchings, J. J., & Burnip, S. R. Birth record assessments and test performance at eight months. American Journal of Diseases of Children, 1965, 109, 416-426.

Houston, Susan H. A re-examination of some assumptions about the language of the disadvantaged child. Child Development, 1970 41, 947-963.

Husén, T. Intra-pair similarities in the school achievements of twins. Scandinavian Journal of Psychology, 1963, 4, 108-114.

Ingle, D. J. Editorial: The need to study biological differences among racial groups: Moral Issues. Perspectives in Biology and Medicine, 1967, 10, 497-499.

Jacobziner, H. Lead poisoning in childhood epidemiology, manifestations and prevention. Clinical Pediatrics, 1966, 5, 277.

Jencks, C. Intelligence and race: What color is IQ? New Republic, September 13, 1969, Nos. 10-11, Issues 2854-2855, pp. 25-29.

Jensen, A. R. Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. Proceedings of the National Academy of Sciences, 1967, 58, 149-156.

Jensen, A. R. Social class, race, and genetics: Implications for education. American Educational Research Journal, 1968, 5, 1-42. (a)

Jensen, A. R. Patterns of mental ability and socioeconomic status. Proceedings of the National Academy of Sciences, 1968, 60, 1330-1337. (b)

Jensen, A. R. Another look at culture-fair tests. In Western Regional Conference on Testing Problems, Proceedings for 1968, "Measurement for Educational Planning." Berkeley, Calif.: Educational Testing Service, Western Office, 1968. pp. 50-104. (c)

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123. (a)

Jensen, A. R. Reducing the heredity-environment uncertainty. Harvard Educational Review, 1969, 39, 449-483. (b)

Jensen, A. R. Criticism or propaganda? American Psychologist, 1969, 24, 1040-1041. (c)

Jensen, A. R. Understanding readiness. Urbana, Illinois: University of Illinois Press, 1969. (d)

Jensen, A. R. IQs of identical twins reared apart. Behavior Genetics, 1970, 1, 133-148. (a)

Jensen, A. R. Hierarchical theories of mental ability. In Dockrell, B. (Ed.) On intelligence. Toronto: Ontario Institute for Studies in Education, 1970. pp. 119-190. (b)

Jensen, A. R. A theory of primary and secondary familial mental retardation. In Ellis, N. R. (Ed.) International Review of Mental Retardation, Vol. 4. New York: Academic Press, 1970. pp. 33-105. (c)

Jensen, A. R. Do schools cheat minority children? Educational Research, 1971, in press. (a)

Jensen, A. R. The race x sex x ability interaction. In Cancro, R. (Ed.) Contributions to intelligence. New York: Grune & Stratton, 1971. (b)

Jensen, A. R. The role of verbal mediation in mental development. Journal of Genetic Psychology, 1971, 118, 39-70. (c)

Jensen, A. R. Selection of minorities in higher education. Toledo Law Review, 1971, Spring-Summer, Nos. 2 & 3, 403-457. (d)

Jensen, A. R., & Rohwer, W. D., Jr. Experimental analysis of learning abilities in culturally disadvantaged children. Final report on OEO Project No. 2404, U. S. Office of Economic Opportunity, 1970.

Jinks, J. L. & Fulker, D. W. Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. Psychological Bulltein, 1970, 73, 311-349.

Jones, H. E. A first study of parent-child resemblances in intelligence. Yearbook of the National Soc. Stud. Educ., 1928, 27 (I), 61-72.

Juel-Nielsen, N. Individual and environment: a psychiatiric-psychological investigation of monozygous twins reared apart. Acta psychiatrica et neurologica Scandinavica, 1965, (Monogr. Suppl. 183).

Kagan, J. S. Inadequate evidence and illogical conclusions. Harvard Educational Review, 1969, 39, 274-277.

Kárpinos, B. D. Qualification of American youths for military service. Medical Statistics Division, Office of the Surgeon General, Department of the Army, 1962.

Kates, Solis L., Kates, W. W., & Michael, J. Cognitive processes in deaf and hearing adolescents and adults. Psychological Monographs 1962, 76, whole No. 551

Katz, I. Review of evidence relating to effects of desegregation on the intelligence performance of Negroes. American Psychologist, 1964, 19, 381-399.

Katz, I. Factors influencing Negro performance in the desegregated school. In Deutsch, M., Katz, I., & Jensen, A. R. (Eds.) Social Class, Race, and Psychological development. New York: Holt, Rinehard, & Winston, 1968. Pp. 254-289.

Kennedy, W. A., Van De Riet, V., and White, J. C., Jr. A normative sample of intelligence and achievement of Negro elementary school children in the Southeastern United States. Monographs of the Society for Research on Child Development, 1963, 28, No. 6.

Kilbride, J. E. The motor development of rural Baganda infants. Unpublished Master's dissertation, Pennsylvania State University, 1969.

Klineberg, O. Negro intelligence and selective migration. New York: Columbia University Press, 1935.

Kline berg, O. (Ed.) Characteristics of the American Negro. New York: Harper, 1944.

Klineberg, O. Race and psychology. In *The Race Question in Modern Science*, UNESCO, 1956.

Klineberg, O. Negro-white differences in intelligence test performance: A new look at an old problem. *American Psychologist*, 1963, 18, 198-203.

Knobloch, Hilda, & Pasamanik, B. Further observations on the behavioral development of Negro children. *Journal of Genetic Psychology*, 1953, 83, 137-157.

Knoblock, H. & Pasamanick, B. The relationship of race and socioeconomic status to the development of motor behavior patterns in infancy. *Psychiatric Research Reports*, 1958, 10, 123-133.

Knoblock, Hilda, & Pasamanick, B. Prospective studies on the epidemiology of reproductive casualty: Methods, findings, and some implications. *Merrill-Palmer Quarterly of Behavior and Development*, 1966, 12, 27-43.

Kohlberg, L. Early education: A cognitive-developmental view. *Child Development*, 1968, 39, 1013-1062.

Kuttner, R. E. Use of accentuated environmental inequalities in research on racial differences. *Mankind Quarterly*, 1968, 8, 147-160.

Langton, E. A. C. Some observations of infants and young persons in Bunyoro, Uganda. *East African Medical Journal*, 1934-35, 11, 316-323.

Laughlin, W. S. Race: A population concept. *Eugenics Quarterly*, 1966, 13, 326-340.

Leahy, Alice M., Nature-nurture and intelligence. *Genetic Psychol. Monographs*, 1935, 17, 241-305.

Lederberg, J. Racial alienation and intelligence. *Harvard Educational Review*, 1969, 39, 611-615.

Lesser, G. S., Fifer, G., & Clark, D. H. Mental abilities of children from different social-class and cultural groups. *Monographs of the Society for Research in Child Development*, 1965, 30, No. 4.

Lewontin, R. C. Race and Intelligence. *Bulletin of the Atomic Scientists*, 1970, 26, No. 3, 2-8. (a)

Lewontin, R. C. Further remarks on race and the genetics of intelligence. *Bulletin of the Atomic Scientists*, 1970, 26, No. 5, 23-25. (b)

Liddicoat, R. Development of Bantu children. *Developmental Medicine and Child Neurology*, 1969, 11, 821.

Liddicoat, R., and Koza, C. Language development in African infants. *Psychologia Africana*, 1963, 10, 108-116.

Light, R. J., & Smith, P. V. Social allocation models of intelligence: A methodological inquiry. *Harvard Educational Review*, 1969, 39, 484-510.

MacArthur, R. Some differential abilities of northern Canadian native youth. International Journal of Psychology, 1968, 3, 43-51.

MacArthur, R. S. Some cognitive abilities of Eskimo, white and Indian-Metis pupils aged 9 to 12 years. Canadian Journal of Behavioral Sciences, 1969, 1, 50-59.

MacArthur, R. S., & Elley, W. B. The reduction of socio-economic bias in intelligence testing. British Journal of Educational Psychology, 1963, 33, 107-119.

Malecot, G. The mathematics of heredity. (trans. by D. M. Yermanos) San Francisco: W. H. Freeman & Co., 1969.

Martin, R., & Saller, K. Lehrbuck der Anthropologie, Vol. II. Stuttgart: Gustav Fischer, 1959.

Massé, G. Croissance et developpement du l'enfant en Dakar. Paris: Centre Internationale du l'Enfance, 1969.

McCall, R. B. Intelligence quotient pattern over age: Comparisons among siblings and parent-child pairs. Science, 1970, 170, 644-648.

McGurk, F. C. J. On white and Negro test performance and socioeconomic factors. Journal of Abnormal and Social Psychology, 1953, 48, 448-450. (a)

McGurk, F. C. J. Socioeconomic status and culturally weighted test scores of Negro subjects. Journal of Applied Psychology, 1953, 37, 276-277. (b)

McGurk, F. C. J. The culture hypothesis and psychological tests. In Kuttner, R. E. (Ed.), Race and modern science. New York: Social Science Press, 1967. pp. 367-381.

McNemar, Q. Twin resemblances in motor skills, and the effect of practice thereon. Journal of Genetic Psychology, 1933, 42, 70-97.

McNemar, Q. Lost: Our intelligence? Why? American Psychologist, 1964, 19, 871-882.

McNemar, Q. A critical examination of the University of Iowa studies of environmental influence upon the IQ. Psychological Bulletin, 1940, 37, 63-92.

Mechanic, D. Medical sociology. New York: The Free Press, 1968.

Miller, J. O., & Phillips, J. A preliminary evaluation of the Head Start and other metropolitan Nashville kindergartens. Unpublished manuscript, Demonstration and Research Center for Early Education, George Peabody College for Teachers, Nashville, 1966.

Moreigne, F., & Senecal, J. Resultats d'un group d'enfants africains au Terman-Merrill. Revue de phychologie appliquee, 1962, 12, 15-32.

Naeye, R. L., Diever, M. M., Dellinger, W. S., & Blanc, W. A., Urban poverty: Effects on prenatal nutrition. Science, 1969, 166, 1026.

Naylor, A. F., & Myrianthopoulos, N. C. The relation of ethnic and selected socioeconomic factors to human birth-weight. Annals of Human Genetics 1967, 31, 71-83.

Nelson, G. K., & Dean, R. F. A. Bull. World Health Organ., 1959, 21, 779. Cited by G. Cravioto, Malnutrition and behavioral development in the preschool child. Pre-school child malnutrition. National Health Science, Public.No. 1282, 1966.

Newman, H. H., Freeman, F. N., & Holzinger, K. J. Twins: A study of heredity and environment. Chicago: University of Chicago Press, 1937.

Nichols, R. C. The National Merit twin study. In Vandenberg, S. G. (Ed.) Methods and goals in human behavior genetics. New York: Academic Press, 1965, pp. 231-244.

Noble, C. E. The learning of psychomotor skills. Annual Review of Psychology, 1968, 19, 203-250.

Noble, C. E. Race, reality, and experimental psychology. Perspectives in Biology and Medicine, 1969, 13, 10-30.

Office of the Surgeon General, Supplement to health of the Army: Results of the examination of youths for military service, 1968. Medical Statistics Agency, Department of the Army, Washington, D. C. June, 1969.

Osborne, R. T. Racial differences in mental growth and school achievement: A longitudinal study. Psychological Reports, 1969, 7, 233-239.

Osborne, R. T. Population pollution. Journal of Psychology, 1970, 76, 187-191.

Osborne, R. T., & Gregor, A. J. The heritability of visualization, perceptual speed, and spatial orientation. Perceptual and Motor Skills, 1966, 23, 379-390.

Osborne, R. T., Gregor, A. J., & Miele, F. Heritability of numerical facility. Perceptual and Motor Skills, 1967, 24, 659-666.

Osborne, R. T., & Gregor, A. J. Racial differences in heritability estimates for tests of spatial ability. Perceptual and Motor Skills, 1968, 27, 735-739.

Pasamanick, B. A comparative study of the behavioral development of Negro infants. Journal of Genetic Psychology, 1949, 75, 82-125.

Pasamanick, B., & Knobloch, Hilda. Retrospective studies on the epidemiology of reproductive casualty: Old and new. Merrill-Palmer Quarterly of Behavior and Development, 1966, 12, 7-26.

Paulson, E. An approximate normalization of the analysis of variance distribution. Annals of Mathematical Statistics, 1942, 13, 233-235.

Peisach, Estelle C. Children's comprehension of teacher and peer speech. Child Development, 1965, 36, 467-480.

Penrose, L. S. Genetics of the human race. In Dunn, L. C. (Ed.), Genetics in the 20th Century. New York: Macmillan Co., 1951.

Peterson, J., & Lanier, L. H. Studies in the comparative abilities of whites and Negroes. Mental Measurements Monographs, 1929, No. 5, Pp. 156.

Phillips, J. L., Jr. The origins of intellect: Piaget's theory. San Francisco: W. H. Freeman, 1969.

Piaget, J. The general problem of the psychobiological development of the child. In J. M. Tanner & B. Inhelder (Eds.), Discussions on Child Development. Vol. 4. New York: International Universities Press, 1960.

Platt, B. S. Nutrition and psychosocial deprivation. In Perspectives on human deprivation: Biological, psychological and sociological. Washington, D. C.: U. S. Department of Health, Education, and Welfare, 1968, pp. 241-245.

Plotkin, L. Negro intelligence and the Jensen hypthesis. The New York Statistician, 1971, 22, 3-7.

Pollitt, E., & Granoff, D. Mental and motor development of Peruvian children treated for severe malnutrition. Revista Interamericana Psicologia, 1967, 1, 93-102.

Quay, Lorene C. The effects of language and reinforcement on the intelligence test performance of Negro children. Child Development, in press.

Race, R., & Sanger, R. Blood groups in man. 5th ed. Philadelphia: Davis, 1968.

Ramarosaona, A. Psychomotor development in early childhood in the Tanararive region. Annex IX in C. S. A. Meeting of Specialists on the Basic Psychology of African and Madasaacan Populations, C. S. A. Publication No. 51, Tanararive, 1959.

Reed, T. E. Caucasian genes in American Negroes. Science, 1969, 165, 762-768. (a)

Reed, T. E. Letters. Science, 1969, 165, 1353. (b)

Rosenfeld, M., & Hilton, T. L. Negro-white differences in adolescent educational growth. American Educational Research Journal, 1971, 8, 267-283.

Rosenthal, R. & Jacobson, Lenore. Pygmalion in the classroom. New York: Holt, Rinehart, & Winston, 1968.

Roth, E. Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. Zeitschrift Fuer Experimentelle Und Angewandte Psychologie, 1964, 11, 616-622.

Sattler, J. M. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Scarr, Sandra. Genetic and environmental variances in aptitude and achievement scores within different racial and social class groups. Unpublished manuscript, 1971.

Schull, W. J., & Neel, J. V. The effects of inbreeding on Japanese children. New York: Harper & Row, 1965.

Scott, R. B., Ferguson, Angella B., Jenkins, M. E., & Cutler, F. E. Growth and development of Negro Infants: V. Neuromuscular patterns of behavior during the first year of life. Pediatrics, 1955, 16, 24-30.

Scrimshaw, N. S., & Gordon, J. E., (Eds.) Malnutrition, learning, and behavior. Cambridge, Mass.: M.I.T. Press, 1968.

Scriven, M. The values of the academy. (Moral issues for American education and educational research arising from the Jensen case). Review of Educational Research, 1970, 40, 541-549.

Sexton, Patricia C. Education and income: Inequalities of opportunity in our public schools. New York: Viking, 1961.

Shields, J. Monozygotic twins brought up apart and brought up together. London: Oxford University Press, 1962.

Shockley, W. Offset analysis description of racial differences. Proceedings of the National Academy of Sciences, 1969, 64, 1432.

Shockley, W. "Cooperative correlation" hypothesis for racial differences in earning power. Proceedings of the National Academy of Sciences, 1970, 66, 245. (a)

Shockley, W. New methodology to reduce the environment-heredity uncertainty about dysgenics. Paper presented at the Autumn meeting of the National Academy of Sciences, Rice University, Houston, Texas, October 21, 1970. (b)

Shockley, W. Negro IQ deficit: Failure of a "Malicious Coincidence" model warrants new research proposals. Review of Educational Research, 1971, 41, 227-248. (a)

Shockley, W. Hardy-Weinberg law generalized to estimate hybrid variance for Negro populations and reduce racial aspects of the environment-heredity uncertainty. Proceedings of the National Academy of Sciences, 1971. (b)

Shuey, Audrey M. The testing of Negro intelligence. (2nd ed.) New York: Social Science Press, 1966.

Sigel, J. E., & Olmsted, Patricia. Modification of cognitive skills among lower-class black children. In Hellmuth, J. (Ed.) Disadvantaged Child, Vol. 3. New York: Brunner-Mazel, 1970, Pp. 300-338.

Sitgreaves, Rosedith. Comments on the "Jensen Report." Paper presented at a meeting of the National Academy of Education, Los Angeles, California, October, 1969.

Sjogren, D. D. Achievement as a function of study time. American Educational Research Journal, 1967, 4, 337-343.

Skeels, H. M. Adult status of children with contrasting early life experiences: A follow-up study. Child Development Monograph, 1966, 31, No. 3, Serial No. 105.

Skeels, H. M., & Dye, H. B. A study of the effects of differential stimulation on mentally retarded children. Proceedings and Addresses of the American Association of Mental Deficiencies, 1939, 44, 114-136.

Skodak, Marie, & Skeels, H. M. A final follow-up study of one hundred adopted children. Journal of Genetic Psychology, 194c, 75, 85-125.

Snow, R. Unfinished Pygmalion. Contemporary Psychology, 1969, 14, 197-199.

Spuhler, J. N., & Lindzey, G. Racial differences in behavior. In. J. Hirsch (Ed.), Behavior-genetic analysis. New York: McGraw-Hill, 1967. Pp. 366-414.

Stanley, J. C. Predicting college success of the educationally disadvantaged. Science, 1971, 171, 640-647.

Stein, Zena A., & Kassab, H. Nutrition. In: Mental retardation. New York: Grune & Stratton, 1970. Pp. 92-116.

Stern, C. Model estimates of the frequency of white and near white segregants in the American Negro. Acta Genetica (Basel), 1953, 4, 281-298.

Stinchcombe, A. L. Environment: the cumulation of effects is yet to be understood. Harvard Educational Review, 1969, 39, 511-522.

Stoch, M. B., & Smythe, P. M. Does undernutrition during infancy inhibit brain growth and subsequent intellectual development? Archives of Diseases of Childhood, 1963, 38, 546-552.

Stouffer, S. A., et al. The American soldier, Vol. I. New York: Wiley, 1965.

Tenopyr, Mary L. Race and socioeconomic status as moderators in predicting machine-shop training success. Paper presented at the annual meeting of the American Psychological Association, Washington, D. C., September 4, 1967.

Terman, L. Genetic studies of genius. Vol. I. Mental and physical traits of a thousand gifted children. Stanford, California: Stanford University Press, 1926.

Terman, L. M. Personal reactions of the Yearbook Committee. In Whipple, G. M. (Ed.). Intelligence: Its nature and nurture, 39th Yearbook of the National Society for the Study of Education. Part I. 1940, Pp. 460-467.

Terman, L. M., & Oden, M. The gifted group at mid-life. Stanford: Stanford University Press, 1959.

Theunissen, K. B. A preliminary comparative study of the development of motor behavior in European & Bantu children up to the age of one year. M. A. Thesis, Natal University College, Durban, 1948.

Thoday, J. M. Limitations to genetic comparison of populations. Journal of Biosocial Science, Supplement, 1969, 1, 1969.

Thoday, J. M. & Gibson, J. B. Environmental and genetical contributions to class difference: a model experiment. Science, 1970, 167, 990-992.

Thorndike, E. L. Human nature and the social order. New York: Macmillan, 1940.

Thorndike, R. L. Community variables as predictors of intelligence and academic achievement. Journal of Educational Psychology, 1951, 42, 321-338.

Thorndike, R. L. Review of R. Rosenthal and L. Jacobson, "Pygmalion in the classroom" American Educational Research Journal, 1968, 5, 708-711.

Tuddenham, R. D. A "Piagetian" test of cognitive development. In Dockrell, B., (Ed.) On intelligence. Toronto: Ontario Institute for Studies in Education, 1970, Pp. 49-70.

Tulkin, S. R. Race, class, family, and school achievement. Journal of Personality and Social Psychology, 1968, 9, 31-37.

Tyler, Leona E. The psychology of human differences. (3rd ed.) Yew York: Appleton-Century-Crofts, 1965.

U. S. Department of Health, Education, and Welfare. Vital statistics of the United States, 1965, Vol. II, Mortality, Part A. Washington, D. C., 1967

Vandenberg, S. G. A comparison of heritability estimates of U. S. Negro and white high school students. Acta Geneticae Medicae et Gemellologiae, 1970, 19, 280-284.

Vernon, M. Relationship of language to the thinking process. Archives of General Psychiatry, 1967, 16, 325-333.

Vernon, M. Fifty years of research on the intelligence of deaf and hard-of-hearing children: A review of literature and discussion of implications. Journal of Rehabilitation of the Deaf, 1968, 1, 1-12.

Vernon, P. E. Ability factors and environmental influences. American Psychologist, 1965, 20, 723-733a.

Vernon, P. E. Environmental handicaps and intellectual development: Part II and Part III. British Journal of Educational Psychology, 1965, 35, 1-22 (b).

Vouilloux, --. Etude de la psycho-motricite d'enfants africaines au Cameroun: test de Gesell et reflexes archaigques. Journal du la Societe des Africanistes, 1959, 29, 11-18.

Voyat, G. IQ: God-given or man-made? Saturday Review, 1969, 52, 74-75. Reprinted in: Hellmuth, J. (Ed.) Disadvantaged child. Vol. 3. New York: Brunner-Mazel, 1970, Pp. 158-162.

Walker, Helen M., & Lev, J. Statistical inference, New York: Holt, 1953.

Waller, J. H. The relationship of fertility, generation length, and social mobility to intelligence test scores, socioeconomic status, and educational attainment. Unpublished Ph.D. dissertation, University of Minnesota, 1969.

Walters, E. Etta. Comparative development of Negro and white infants. Journal of Genetic Psychology, 1967, 110, 243-251.

Warren, B. L. A multiple variable approach to the assortative mating phenomenon. Eugenics Quarterly, 1966, 13, 285-290.

Watson, P. How race effects IQ. New Society, 1970, 16, 103-104.

Wechsler, D. The measurement and appraisal of adult intelligence. 4th Edition. Baltimore: Williams & Wilkins, 1958.

Weener, P. D. Social dialect differences and the recall of verbal messages. Journal of Educational Psychology. 1969, 60, 194-199.

Weyl, N. Some comparative performance indexes of American ethnic minorities. Mankind Quarterly, 1969, 9, 106-128.

White, S. H. Evidence for a hierarchical arrangement of learning processes. In L. P. Lipsitt & C. C. Spiker (Eds.) Advances in child development and behavior. Vol. 2. New York: Academic Press, 1965. Pp. 187-220.

Williams, Judith R., & Scott, R. B. Growth and development of Negro infants: IV. Motor development and its relationship to child-rearing practices in two groups of Negro infants. Child Development, 1953, 24, 103-121.

Wilson, A. B. Educational consequences of segregation in a California community. Racial isolation in the public schools, Appendices, Vol. II, p. 185. U. S. Commission on Civil Rights, Washington, 1967.

Winick, M. Nutrition, growth, and mental development: Biological correlations. American Journal of Diseases of Children, 1970, 120, 416-418.

Wolański, N., Jarosz, Emilia, & Pyłuk, Mira. Heterosis in man: Growth in offspring and distance between parents' birthplaces. Social Biology, 1970, 17, 1 - 16.

Wright, S. Statistical methods in biology. Journal of the American Statistical Association, 1931, 26, 155-163.

Young, M., and Gibson, J. B. Social mobility and fertility. In Biological aspects of social problems, J. E. Meade and A. S. Parkes (Eds.), Edinburgh: Oliver and Boyd, 1965.

Zirkel, P. A., & Moses, E. G. Self-concept and ethnic group membership among public school students. American Educational Research Journal, 1971, 8, 253-265.

STATEMENT OF JANE R. MERCER

I—INTRODUCTION

Mr. Chairman and members of the Committee. My name is Jane R. Mercer and I am an Associate Professor of Sociology at the University of California, Riverside and a Research Specialist with the California Department of Mental Hygiene. I hold an M.A. degree from the University of Chicago and a Ph. D. degree in Sociology from the University of Southern California. During the past ten years, I have been directing a study of mental retardation in the city of Riverside, California, population 130,000, under grants from the National Institute of Mental Health and have been studying school desegregation under grants from the California State Department of Education, Division of Compensatory Education and the Office of Education, DHEW, Title IV. A comprehensive report of the study of mental retardation is being published by the University of California Press under the title *Labeling The Mentally Retarded* and preliminary reports of the findings from this study had been made to the President's Committee on Mental Retardation and have appeared in various publications, among them "Institutional Anglocentism: Labeling Mental Retardates in the Public Schools" in *Race, Change, and Urban Society*, Volume 5, Urban Affairs Annual Reviews edited by Peter Orleans and William R. Ellis, Sage Publications, 1971 and "Sociocultural Factors in Labeling Mental Retardates" in the *Peabody Journal of Education*, Volume 48, No. 3, April, 1971.

II—THE PUBLIC SCHOOL'S ROLE IN LABELING THE MENTALLY RETARDED

We did a study of all the persons labeled as mentally retarded by contacting 241 different organizations in Riverside and asking each organization to give us information on each mentally retarded person being served by that group. The public schools nominated 429 of the 812 persons on the case register and 340 of them were not nominated by any other organization.

We found that the public schools rely more on IQ test scores than any other community agency. Ninety-nine percent of the persons nominated by the schools had been given an IQ test but only 13% had received a medical diagnosis. We found that 46% of the persons nominated as mental retardates by the public schools had IQs above 70 and 62% had no reported physical disabilities. All other community agencies, except law enforcement, were labeling persons with significantly lower IQ test scores and more physical disabilities.

In California, the rates for placing Mexican-American and Black children in classes for the mentally retarded are two to four times

higher per thousand than the rates for English-speaking Caucasian children, whom I will henceforth call Anglos. In Riverside, we found that ethnic disproportions were especially marked among public schools nominees. There were 4½ times more Mexican-American children and twice as many Black children in classes for the mentally retarded as would be expected from their proportion in the population. On the other hand, there were only half as many Anglo children in these classes as we would expect from their proportion in the population.

When we studied the labeling process in the public schools, we found that ethnic disproportions first appeared in the labeling process at the point when the IQ test was administered. We also found that the Black and Mexican-American children who were placed in special education classes had higher IQ test scores and fewer physical disabilities than the Anglo children placed in those classes. We found no evidence in our study that these ethnic disproportions resulted from a conscious policy of discrimination. However, there is no doubt that the labeling process is Anglocentric and weighs most heavily on persons from lower socioeconomic statuses and minority ethnic groups.

III—Conclusions From the Community Survey of Mental Retardation

These findings lead us to concentrate our efforts on identifying which aspects of the clinical assessment process are producing ethnic disproportions. We studied a representative sample of 6,907 persons from the general population of the community. We reached three major conclusions in this portion of our study. We found that the majority of the adults with IQs between 70 and 85 were, in fact, filling the usual complement of social roles for persons of their age and sex. Their role performance appeared neither subnormal nor particularly unusual. We concluded that the 3% cutoff, that is, an IQ below 70, was the criterion most likely to identify persons in need of special assistance and supervision and least likely to stigmatize as mentally retarded persons who would be filling a normal complement of social roles as adults. We concluded that the public schools should use a cutoff level of IQ 69 and below rather than their customary cutoff of IQ 79 and below in defining the mentally retarded.

The definition of mental retardation proposed by the American Association for Mental Deficiency requires subnormality in both intellectual performance and adaptive behavior but most psychologists give only an IQ test when making assessments. When we secured information about the adaptive behavior of individuals in our sample who had low IQ test scores, we found that 80% of the adults had graduated from high school; all had held jobs; 65% had white collar positions.

Their social role performance tended to be indistinguishable from that of other adults in the community. We found that 60% of the Mexican-Americans and 91% of the Blacks who had IQ test scores below 70 passed the adaptive behavior measure while none of the Anglos with IQs this low were performing normally in their social roles. We concluded that the IQ test is not as valid a predictor of social

role performance for Mexican Americans and Blacks as for Anglos. Thus, the evaluation of adaptive behavior was especially important in assessing persons from ethnic minorities and lower socioeconomic levels, persons from backgrounds that do not conform to the average sociocultural pattern of American society. Many of them may fail IQ tests mainly because they have not had the opportunity to learn the cognitive skills and to acquire the knowledge needed to pass such tests. They demonstrate by their ability to cope with problems in other areas of life that they are not comprehensively retarded. We concluded that a child ought to fail both criteria before being labeled as mentally retarded.

Our third major conclusion dealt with cultural biases in IQ tests. The IQ tests now being used by psychologists are, to a large extent, Anglocentric. These tests tend to measure the extent to which a child's family background is similar to that of the middle class Anglo-American core culture. We found that approximately 32% of the differences in IQ tests scores among a sample of approximately 1,500 Black, Mexican-American, and Anglo elementary school children in one California school district could be accounted for by differences in the sociocultural characteristics of their families. We concluded that sociocultural factors should be taken into account when interpreting the meaning of any child's IQ test score.

To do this, we grouped each Black and Mexican-American elementary school child in our sample into one of five groups according to the extent to which his family background conformed to the average configuration for the total community. Each child was given one point for each family characteristic which was like the dominant society on the five most important sociocultural variables which we found were correlated with Full Scale WISC IQ for his ethnic group. If his family was similar to the dominant society on all five characteristics he received a score of five. If his background was similar to the dominant society on four characteristics he received a score of four, and so forth.

The average IQ for the entire group of Mexican-American children was 90.4%. The 127 children from backgrounds least like the dominant society, those having zero or 1 modal characteristic, had an average IQ of 84.5. The 146 children with two modal background characteristics had a mean IQ of [illegible].1; those with three characteristics a mean IQ of 89.0; those with four characteristics a mean IQ of 95.5; and those with all five modal characteristics had a mean IQ of 104.4. When social background was held constant there was no difference between the measured intelligence of Mexican-American children and the Anglo children on whom the test was standardized.

The total group of 339 Black children had an average IQ of 90.5 when there was no control for sociocultural factors. The 47 children who came from backgrounds least like the dominant community had an average IQ of 82.7. Those with two modal characteristics had an average IQ of 87.1. Those with three characteristics had an IQ of 92.8; those with four characteristics had an IQ of 95.5; and those with five characteristics an IQ of 99.5, exactly at the national norm for the test. Thus, Black children who came from family backgrounds comparable to those of the middle class Anglo community did just as well on the Weschler Intelligence Scale for Children as the children on whom the test was standardized. When sociocultural differences were held constant, there were no differences in measured intelligence.

We concluded on the basis of our study, that diagnostic procedures in the public schools should be broadened to reflect the pluralistic nature of American society. We are proposing the development of pluralistic assessment procedures which involve securing information beyond that ordinarily considered in public school assessment. Our findings suggest that only persons in the lowest 3% of the population, that is with IQs under 70, should be labeled as comprehensively retarded. Our findings also suggest that information about adaptive behavior—a child's ability to cope with problems in the family, neighborhood, and community—should be considered as well as his IQ test score when making a clinical assessment. Only persons who are subnormal *both* on the IQ test and in adaptive behavior should be regarded as comprehensively retarded. Finally, in pluralistic assessment, the meaning of a particular IQ test score or adaptive behavior score should be interpreted not only within the framework of the standardized norms based on a sample of Anglo children but should also be evaluated in relation to the norm for the sociocultural group to which the child belongs. His position on the standard norms indicates how well he is likely to do in a regular public school classroom without special assistance. His position on the norms for his own sociocultural group indicates his probable potential for learning.

When we reanalyzed the data from our survey of mental retardation in the community using pluralistic diagnostic procedures, ethnic differences in rates for mental retardation disappeared. Approximately the same percentage of persons in each ethnic group were identified as comprehensively retarded. We re-evaluated the 268 children who were enrolled in classes for the educable mentally retarded in two Southern California school districts using pluralistic diagnostic procedures. We found that approximately 75% of the children in those classes would not have been labeled as comprehensively retarded if their adaptive behavior and their sociocultural backgrounds had been taken into account at the time they were evaluated.

IV—Implications for Equal Educational Opportunity

These findings bring us to the broader issue of the educational responsibility of the public schools to children of all racial and cultural groups in American society. In the United States, the cultural pattern of that segment of the population consisting of white, Anglo-Saxon Protestants has had the primary influence on the educational system. Consequently, the public schools have, in the past, seen their primary task as perpetuating the language, literature, values, and history of the Anglo segment of the population. The content of the standardized American IQ tests reflect this cultural orientation. Children who have not had the opportunity to acquire the skills and learn the knowledge covered in such tests are severely penalized when these scores are misinterpreted. This discrimination becomes particularly invidious when the IQ test scores of children of minority ethnic and cultural groups are viewed, primarily, as a measure of inherited mental ability and these children are systematically assigned to public school programs and classes which divert them from the mainstream of public education.

Because the schools are the primary social institution allocating persons to adult roles and statuses in American society, the kind and amount of education which a person has determines, to a large extent, whether he will participate in the mainstream of American life or be shunted into the byways. Educational decisions based on tests which systematically favor one group over another predetermine which group will occupy the seats of power and which group will remain powerless. Thus, the practice of interpreting scores on an IQ test as primarily the result of inherited characteristics has far reaching implications for equal educational opportunity.

Biological potential cannot be measured directly. To compare the average IQ test score of one group with the average IQ test score of another in order to estimate their relative biological potential requires that at least five conditions be met. Such comparisons can be made only if there is (1) equal exposure to opportunities to learn the culture-bound materials in the test, (2) equal participation in the values and motivational systems assumed by the test, (3) equal freedom from anxiety and emotional disturbance, (4) equal familiarity with and comfort in the test situation, (5) equal freedom from environmentally produced physical disabilities that might interfere with performance. This situation clearly does not hold true when comparing the average performance of majority children in the United States with that of minority children.

In a sample of 180 Anglo, 180 Mexican-American, and 180 Black elementary school children, we did a series of partial correlations in which we correlated ethnic group with IQ test score and held four sociocultural conditions constant: socioeconomic status; the extent to which the child's mother valued individualistic achievement; the extent to which the child's mother participated in formal community organizations; and the extent to which the neighborhood in which the child lived was desegregated. The correlation between ethnic group and IQ was reduced essentially to zero when these sociocultural differences were held constant. This indicates that differences in performance of the children on this IQ test could be accounted for, statistically, by these sociocultural differences. These findings would indicate that educational programs should be planned on the assumption that children from all ethnic, racial, and cultural groups in American society have essentially the same ability to learn and that present differences in average group performance are the result of lack of exposure of minority children to the mainstream of American society. Every effort should be made through educational programming, school desegregation, and neighborhood desegregation to expand the amount and quality of contact between the various ethnic and cultural groups that make up America and to develop a public school program that more nearly represents the rich cultural variety of American life.

SOCIOCULTURAL FACTORS IN THE EDUCATIONAL EVALUATION OF BLACK AND CHICANO CHILDREN*

BY JANE R. MERCER, ASSOCIATE PROFESSOR, SOCIOLOGY
UNIVERSITY OF CALIFORNIA, RIVERSIDE

Data in this report has been collected under the auspices of the following grants: Public Health Service Research Grant No. MH-08667, from the National Institute of Mental Health, Department of Health, Education, and Welfare and Public Health Service General Research Support Grant No. 1-S01-FR-05632-02, from the Department of Health, Education, and Welfare, Sociobehavioral Study Center in Mental Retardation, Pacific State Hospital, Pomona, Calif.; Public Health Service Grant No. PH43-67-756; McAteer Grant No. M8-14A and M9-14 from the California State Department of Education, Office of Compensatory Education.

We wish to thank key persons from the Riverside Unified School District for their support and advice in our research efforts: Past superintendent of schools, Bruce Miller; superintendent of schools, Raymond E. Berry; Albert Marley, director of pupil personnel services; former director of pupil personnel, Richard Robbins; Mabel C. Purl, director of research and testing. From the Alvord School District we acknowledge with gratitude the help of Robert W. Hocker, director of pupil personnel; and J. Martin Kaeppel, psychometrist.

AUTHOR'S NOTE.—A more comprehensive and detailed analysis and report of the referral process, the clinical testing process, the assumptions and inferences of the clinician, and the pluralistic evaluation process discussed herein appear in a forthcoming volume entitled "Labeling the Mentally Retarded."

The purpose of this conference is to examine the question of human rights in education. The importance of recognizing the rights of each child extends far beyond the school years. The schools are the primary social institution allocating persons to adult statuses and roles in American society. The kind and amount of education which a person has determines, to a large extent, whether he will participate in the mainstream of American life or be shunted into the byways. Educational decisions which systematically favor one group over another predetermine which group will occupy the seats of power and which group will remain powerless. It is this awesome responsibility which I would like to discuss with you this evening.

Although my research studies have been primarily in the area of mental retardation and the findings I will present this evening are based on those studies, I believe these findings have implications be-

*Presented at the 10th Annual Conference on Civil and Human Rights of Educators and Students, Washington, D.C., National Education Association, February 18-20, 1972.

yond the labeling and placement of children in classes for the mentally retarded in the public schools.

In a recent study, we interviewed the mothers of 268 children who were in classes for the educable mentally retarded in two public school districts in southern California. The responses of some of these mothers illustrates, graphically, the three issues which I would like to address tonight: biases in the assessment procedures used to label children as mentally retarded; the stigmatization associated with special class placement; and inadequate programing. I will discuss each of these issues separately and will then present what appears to me to be viable alternatives to present procedures.

We turn first to the issue of systemic cultural biases in the diagnostic procedures used to label children as mentally retarded. Studies dating back to the 1930s have repeatedly demonstrated the cultural biases inherent in IQ tests and other standardized achievement measures. Yet, in spite of these studies, clinicians have continued to interpret children's performances on these tests as if there were no cultural biases and have never systematically taken sociocultural differences into account when interpreting the meaning of a particular child's score. Consequently, we find many children in classes for the mentally retarded whose adaptive behavior, in nonacademic settings, clearly demonstrates that their problems are school specific and that they are not comprehensively incompetent.

John is a 16-year-old, black boy who has been in classes for the educable mentally retarded for the past 8 years. He has an IQ of 83. When John's mother was asked to describe what he does on Saturdays and around the house and in the neighborhood, she gave the following reply:

> John works on weekends at the service station as an attendant * * * I would say he's a good mechanic. He likes to work on cars, changes the oil, helps with overhauling a car, and works on motorcycles. Sometimes he irons, washes dishes mops the floor, cuts the lawn, sweeps off the driveway, goes to the store, runs errands, vacuums the rugs, makes his own bed, and things like that. He's never still too much. He likes to be outdoors and likes to ride motorcycles. He plays basketball and football on Saturdays, works on cars, and then goes to bed.

Pete is a 14-year-old, chicano boy with an IQ of 79 who has just been returned to regular classes after being in special education since he was 10 years old. His mother described him as follows:

> Pete is a very bright child, he's always thinking and doing something; he built a two-room tree house that is just beautiful. He is good at anything that needs putting together. He makes cars with motors, he makes them from old boxes, tires, wood, anything he can find. He makes cages for the animals * * * He is very helping. Sometimes, when I am trying to do some plumbing but cannot do it, he knows how to fix things around here. He is good at plumbing and at figuring things out.

Disproportionately large numbers of black and chicano children are labeled as mentally retarded by the public schools. This phenomenon appears to be true throughout the United States. For example, in California, the rates for placing chicano and black children in classes for the mentally retarded are two to four times higher per thousand than the rates for English-speaking Caucasian children, whom I will henceforth call Anglos.

We did a study of all the persons labeled as mentally retarded in a city of 100,000 persons in southern California. We contacted 241 different organizations in the community and asked each organization to give us information on each mentally retarded person being served by that group. The public schools nominated 429 of the 812 persons on the case register and 340 of them had not been nominated by any other organization. When we studied the number of persons jointly nominated by more than one type of organization, the public schools clearly held the commanding position. They not only labeled more persons as mentally retarded than any other organizations but they shared their labels more widely throughout the community.

We found that the public schools rely more on IQ test scores than any other community agency. Ninety-nine percent of the persons nominated by the schools had been given an IQ test but only 13 percent had received a medical diagnosis. We found that 46 percent of the persons nominated as mental retardates by the public schools had IQs above 70 and 62 percent had no reported physical disabilities. All other community agencies, except law enforcement were labeling persons with significantly lower IQ test scores and more physical disabilities. We concluded that the public school system is the primary labeler in the community. The schools label more persons as mentally retarded, share their labels with more other organizations, and label more persons with IQs above 70 and with no physical disabilities than any other formal organization in the community.

School age children were "over labeled" and preschool children and adults were "under labeled" compared to their percentage in the general population of the community. Before children get to school, only those with the most physical disabilities and the lowest IQs are identified. After graduation from school, most of the persons labeled as mentally retarded in the public schools disappear into the general population and are no longer so labeled. Only the most intellectually and physically subnormal adults continue to be regarded as mental retardates.

We found that ethnic disproportions were especially marked among public school nominees. There were 4½ times more chicano children and twice as many black children in classes for the mentally retarded as would be expected from their proportion in the population. On the other hand, there were only half as many Anglo children in these classes as we would expect from their proportion in the population. When we studied the labeling process in the public schools, we found that teachers and principals were not referring disproportionately large numbers of minority children for psychological evaluation. Ethnic disproportions first appeared in the labeling process at the point

when the IQ test was administered. We also found that the black and chicano children who were placed in special education classes had higher IQ test scores and fewer physical disabilities than the Anglo children placed in those classes.

What produces these differences? Some minority parents were convinced that the special education program was deliberately planned to keep minority children from receiving a full education. According to one black mother whose 14-year-old son had been in special education classes for 5 years "this program is a conspiracy to keep the minorities down. They put as many as possible in these classes because it means more money for the schools. Many times it's because of racial prejudice or behavior problems. I do know, it's most unfair." We found no evidence in our study that these ethnic disproportions resulted from a conscious policy of discrimination. However, there is no doubt that the labeling process is Anglocentric and weighs most heavily on persons from lower socioeconomic statuses and minority ethnic groups.

These findings lead us to concentrate our efforts on identifying which aspects of the clinical assessment process are producing ethnic disproportions. We studied a representative sample of 6,907 persons from the general population of the community, using the American Association for Mental Deficiency definition for mental retardation: Mental retardation refers to a person who is subaverage both in general intellectual functioning and adaptive behavior. This is a two-dimensional definition with two primary symptoms: subnormality in intellectual performance and subnormality in adaptive behavior. Combinations of these two dimensions produce the four major types of persons shown in table 1. The comprehensively retarded are those who are subnormal in both IQ and adaptive behavior. The quasi-retarded are those who are subnormal in IQ but normal in adaptive behavior. The behaviorally maladjusted are those who have normal IQs but are subnormal in adaptive behavior while the normals are those who pass both dimensions. We are concerned primarily with two categories in this typology, the comprehensively retarded and the quasi-retarded.

TABLE 1.—TYPOLOGY OF MENTAL RETARDATION

	Intellectual performance	Adaptive behavior
Comprehensibly Retarded	Subnormal	Subnormal.
Quasi-retarded	Subnormal	Normal.
Behaviorally maladjusted	Normal	Subnormal.
Normals	Normal	Normal.

Intellectual adequacy was measured using standardized measures of intelligence, primarily the Stanford-Binet LM and the Kuhlman-Binet. Because there are no generally accepted measures of adaptive behavior, we developed a series of 28 age-graded scales for this purpose. We conceptualized adaptive behavior as an individual's ability to play ever more complex roles in a progressively widening circle of social systems.

We reached three major conclusions in this portion of our study. Our first finding was that the IQ cutoff used by educational institutions in defining mental retardation is one factor producing ethnic disproportions in the labeling process. Three cutoff levels are currently used for defining subnormality—the American Association of Mental Deficiency defines "subnormal" as performance on a standard measure of intellectual functioning which is greater than one standard deviation below the population mean, approximately the lowest 16 percent of the population (Heber, 1961). Educational practice generally places the dividing line somewhat lower. The highest IQ test score for placement in a class for the educable mentally retarded ranges between 75 and 79, depending upon local usage. This cutoff includes approximately the lowest 9 percent of the population. The test designers suggest a cutoff that more closely conforms with traditional definitions, an IQ below 70, approximately 3 percent of the population (Wechsler, 1958; Termana & Merrill, 1960).

We found that the majority of the adults with IQs between 70 and 85 were, in fact, filling the usual complement of social roles for persons of their age and sex: 84 percent had completed eight grades or more in school; 83 percent had held a job, 65 percent had a semiskilled or higher occupation, 80 percent were financially independent or a housewife, almost 100 percent were able to do their own shopping and to travel alone, and so forth. It is clear that most adults who appear in the borderline category were managing their own affairs and did not appear to require supervision, control, and care for their own welfare. Their role performance appeared neither subnormal nor particularly unusual.

We also found that proportionately more low status persons and persons from minority ethnic groups were defined as comprehensively retarded as the cutoff level for subnormality was raised. When the traditional definition of IQ 69 or below was used, ethnic disproportions were greatly reduced. We concluded that the 3-percent cutoff; that is, an IQ below 70, was the criterion most likely to identify persons in need of special assistance and supervision and least likely to stigmatize mentally retarded persons who would be filling a normal complement of social roles as adults. We concluded that persons scoring in the so called borderline category should be regarded as low normals rather than as comprehensively retarded.

Our second finding concerned the two-dimensional definition of mental retardation proposed by the American Association for Mental Deficiency. Although this definition requires subnormality in both intellectual performance and adaptive behavior, in actual clinical practice, most psychologists give only an IQ test when making assessments. Would it make a difference if psychologists also evaluated adaptive behavior?

We compared the social role performance of the quasi-retarded; that is, those who failed only the IQ test, with the comprehensively retarded; that is, those who failed both the IQ test and the adaptive behavior scales. We found that most quasi-retarded school age children, in spite of their low IQ test score, had avoided falling behind their age mates or being placed in special programs. We found that 80 percent

of the quasi-retarded adults had graduated from high school; they all read books, magazines, and newspapers; all had held jobs; 65 percent had white collar positions. All of them were able to work without supervision; participated in sports; traveled alone; went to the store by themselves; and participated in informal visiting with coworkers, friends, and neighbors. In other words, their social role performance tended to be indistinguishable from that of other adults in the community.

We found that a large percentage of persons in the quasi-retarded category were chicano and black. We found that 60 percent of the chicanos and 91 percent of the blacks who had IQ test scores below 70 passed the adaptive behavior measure while none of the Anglos with IQs this low were performing normally in their social roles. The IQ test is not as valid a predictor of social role performance for chicanos and blacks as for Anglos. Thus the evaluation of adaptive behavior was especially important in assessing persons from ethnic minorities and lower socioeconomic levels, persons from backgrounds that do not conform to the average sociocultural pattern of the community. Many of them may fail IQ tests mainly because they have not had the opportunity to learn the cognitive skills and to acquire the knowledge needed to pass such tests. They demonstrate that they are not comprehensively incompetent by their ability to cope with problems in other areas of life. We concluded that the schools should adhere to the AAMD definition of mental retardation and should develop a systematic method for measuring adaptive behavior as well as IQ in making psychological assessments. We concluded that a child ought to fail both criteria before being labeled as mentally retarded. When we followed this procedure, ethnic disproportions were reduced but still were not completely eliminated.

Our third major conclusion dealt with cultural biases in IQ tests. The IQ tests now being used by psychologists are, to a large extent, Anglocentric. These tests tend to measure the extent to which a child's family background is similar to that of the middle class Anglo-American core culture. We found that approximately 32 percent of the differences in IQ test scores among a sample of approximately 1,500 black, chicano, and Anglo elementary schoolchildren in one California school district could be accounted for by differences in the sociocultural characteristics of their families. We concluded that sociocultural factors should be taken into account when interpreting the meaning of any child's IQ test score.

To do this, we grouped each black and chicano elementary school child in our sample into one of five groups according to the extent to which his family background conformed to the average configuration for the total community. Each child was given one point for each family characteristic which was like the dominant society on the five most important sociocultural variables which we found were correlated with full scale WISC IQ for his ethnic group. If his family was similar to the dominant society on all five characteristics he received a score of five. If his background was similar to the dominant society on four characteristics he received a score of four, and so forth.

Convergence of the Average IQ Test Scores of Chicano Children with the Standard Norms as Sociocultural Factors are Increasingly Controlled

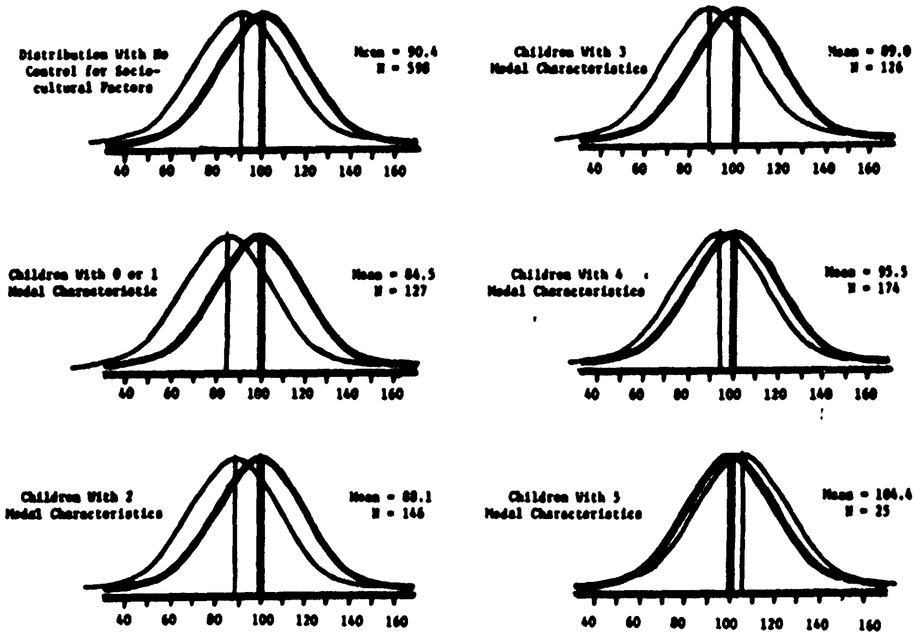


FIGURE 1

The drawings in figure 1 depict the IQ scores of the chicano children in the five sociocultural groupings and compare them with the distribution of IQ scores for the Anglo children on whom the test was standardized. The average IQ for the entire group of chicano children was 90.4. The 127 children from backgrounds least like the dominant society, those having zero or one modal characteristic, had an average IQ of 84.5, borderline mentally retarded by the American Association for Mental Deficiency definition. The 146 children with two modal background characteristics had a mean IQ of 88.1; those with three characteristics a mean IQ of 89; those with four characteristics a mean IQ of 95.5; and those with all five modal characteristics has a mean IQ of 104.4. When social background was held constant there was no difference between the measured intelligence of Mexican-American children and the Anglo children on whom the test was standardized.

Figure 2 shows that the situation is just as dramatic for black children. The total group of 339 black children had an average IQ of 90.5 when there was no control for sociocultural factors. The 47 children who came from backgrounds least like the dominant community had an average IQ of 82.7. Those with two modal characteristics had an average IQ of 87.1. Those with three characteristics had an IQ of 92.8; those with four characteristics had an IQ of 95.5; and those with five characteristics an IQ of 99.5, exactly at the national norm for the test. Thus, black children who came from family backgrounds comparable to those of the middle class Anglo community did just as well on the Weschler Intelligence Scale for Children as the children on whom the test was standardized. When sociocultural differences were held constant, there were no differences in measured intelligence.

Convergence of the Average IQ Test Scores of Black Children with the Standard Norms as Sociocultural Factors are Increasingly Controlled

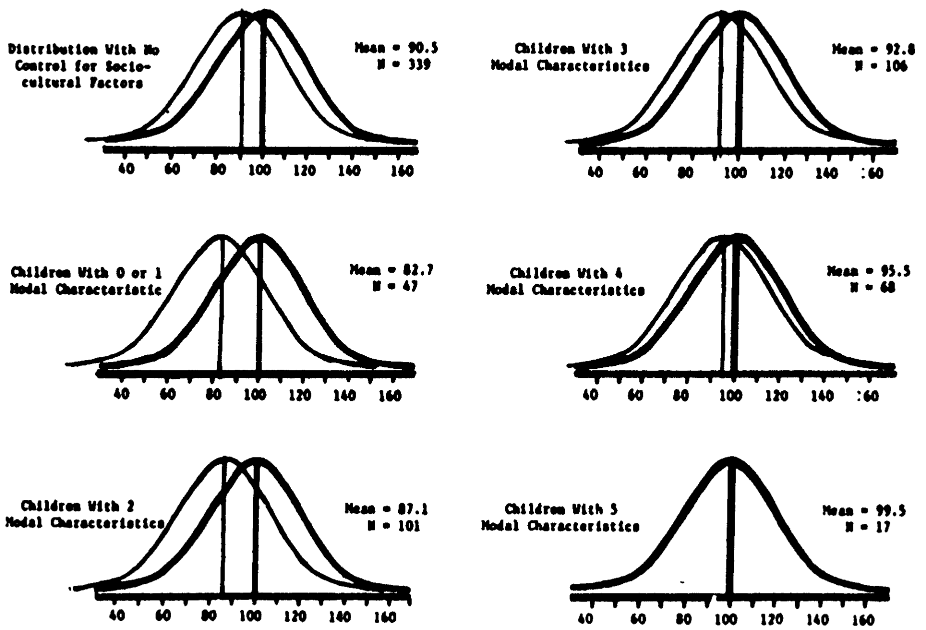


FIGURE 2

We concluded on the basis of our study, that diagnostic procedures in the public schools should be broadened to reflect the pluralistic nature of American society. We are proposing the development of pluralistic assessment procedures which involve securing information beyond that ordinarily considered in public school assessment. Our findings suggest that only persons in the lowest 3 percent of the population; that is, with IQs under 70, should be labeled as comprehensively retarded. Our findings also suggest that information about adaptive behavior—a child's ability to cope with problems in the family, neighborhood, and community—should be considered as well as his IQ test score when making a clinical assessment. Only persons who are subnormal *both* on the IQ test and in adaptive behavior should be regarded as comprehensively retarded. Finally, in pluralistic assessment, the meaning of a particular IQ test score or adaptive behavior score should be interpreted not only within the framework of the standardized norms based on a sample of Anglo children but should also be evaluated in relation to the norm for the sociocultural group to which the child belongs. His position on the standard norms indicate how well he is likely to do in a regular public school classroom with no special assistance. His position on the norms for his own sociocultural group indicates his probable potential for learning.

When we reanalyzed the data from our survey of mental retardation in the community using pluralistic diagnostic procedures, ethnic differences in rates for mental retardation disappeared. Approximately the same percentage of persons in each ethnic group were identified as comprehensively retarded. We reevaluated the 268 children who were enrolled in classes for the educable mentally retarded in two southern California school districts using pluralistic diagnostic pro-

cedures. We found that approximately 75 percent of the children in those classes would not have been labeled as comprehensively retarded if their adaptive behavior and their sociocultural backgrounds had been taken into account at the time they were evaluated.

TABLE 2.—RECLASSIFICATION RESULTING FROM THE APPLICATION OF PLURALISTIC ASSESSMENT PROCEDURES IN REEVALUATING 268 LABELED RETARDATES IN THE PUBLIC SCHOOLS

Ethnic distribution	Handicapped mental retardates: Physically	Handicapped mental retardates: Non-physically	Quasi-retarded	Behaviorally maladjusted	Low normals	Total
Number	32	37	39	78	82	268
Percent of Anglo	84.3	70.2	61.5	41.0	32.9	54.0
Percent of chicano	6.2	16.2	17.9	20.5	40.2	23.0
Percent of black	9.3	13.5	20.5	38.4	26.8	23.0

Table 2 shows the percentage of each ethnic group in each category before and after reevaluation. Although 75 percent of the children in the two school districts were Anglo, only 50 percent of the children in classes for the educable retarded were Anglo. Approximately 25 percent were chicano and 25 percent black. When sociocultural characteristics were taken into account, the distribution by ethnic groups in the category of the mentally retarded closely approximated the distribution for the population of the two school districts. More blacks and chicano children were reclassified as quasi-retarded, behaviorally maladjusted, or low normal. Significant numbers of Anglo children were also reclassified into those categories. Based on this experience, we are proposing that pluralistic assessment procedures be developed and be used systematically in the evaluation of children from nonmodal sociocultural backgrounds.

We turn now to the second major issue to be discussed this evening—that of the stigmatization associated with present labeling and placement practices. Stigmatization was a major concern of parents we interviewed. Many freely expressed their feelings about the special classes and their distress at the psychological consequences for their children of being placed in classes for the educable mentally retarded. One black mother told our interviewer:

> I feel that if a child is put in a special education class in elementary school, by the time he gets to junior high he should be removed because there is a stigma that goes with a special class. Let's face it, children can be real cruel. I feel for the most part the youngsters that are in those classes and retained suffer a great emotional handicap. It's as if they have a sign around their necks for everyone to read "Bill is being retarded in special education." He doesn't like being labeled as retarded. It's affecting him. He begs us to have him removed from that class.
>
> We have to make Bill go to school because that class does not offer a challenge to him. What they do is repetitious—the same thing over and over. * * * He does not like school. We have to make him go. The only reason he consents to go is because we have been promised that he'll be taken out of that EMR class. The teachers have asked us to let them put another one of our kids in EMR. We said an emphatic "No!" because we knew what it was all about.

Maria is a 13-year-old chicano girl who has been in special education classes ever since she was 8 years old. She has a full scale IQ of 62. Her mother told the interviewer that Maria does not want to go to school and goes up to her room and hides until the schoolbus goes away. She knows that her mother cannot take her because her mother does not drive. Maria's mother said that she was very sorry she had signed the papers and that she is not going to sign the paper for the other boys that the school wants to put in special education. She did it once but is not going to do it again.

Parents reported that their children were ashamed to be seen entering the "MR" room because they were often teased by other children about being "MR." The children dreaded receiving mail that might bear compromising identification. They could not understand why they were classified with Anglo children who were physically handicapped when they had no physical disabilities.

Parents were also concerned with the third issue which I would like to discuss this evening—the quality of the educational program in the self-contained special education class. Parents asked why their children were not taught to read like they would be taught in the regular classes. Many parents saw the program as a "sentence of death." We found several parents, like the parents quoted earlier, strenuously resisting the efforts of the school to place younger children in special education classes because they had found it to be an inescapable dead-end for their older children. Our findings confirmed their suspicions. We followed a group of 108 children for several years. Only one child in five ever returned to the regular class. The remaining children either aged out of the program, dropped out of school, or were sent to other special programs or institutions.

During the past 20 years, one of the great achievements of public education in America has been the development of special education programs designed to meet the special needs of handicapped children. It would be a tragedy if these valuable programs were to be jeopardized because of inadequacies in assessment procedures and programing. I believe there are viable alternatives to present practices. Assessment procedures can be modified to take sociocultural factors into account and programing can be altered to reduce stigmatization and to keep many children who are now in self-contained special education classrooms in the educational mainstream.

First, I believe school psychologists should be required to enlarge the scope of information they use in making educational decisions. They should regularly and systematically secure information about the child's adaptive behavior in nonschool situations—at home, in the neighborhood, and in the community. If a child is performing adequately in these settings, then it is clear that his problems are school specific and that he is not comprehensively retarded. His program should be planned with the expectation that he will probably be able to fill his adult roles acceptably and that his primary needs are for special help with academic tasks. For him, special tutoring, programed learning, cross-age teaching, remedial reading, or similar programs are to be preferred to the self-contained classroom and a curriculum for the mentally retarded.

In addition to adaptive behavior, I believe that school psychologists should be required to secure systematic information about a child's sociocultural background which can be used in interpreting the mean-

ing of his IQ test score. Pluralistic norms should be developed so that a child's performance can be compared not only with the performance of the general population, which is composed primarily of Anglo children, but can be compared with the performance of other children from his own sociocultural background. Children from comparable backgrounds would have had similar opportunities to acquire the skills and knowledge covered in the tests. Thus, children whose low performance is primarily the result of sociocultural differences, would be identified and could receive appropriate educational assistance.

I do not agree with those who say we must stop all educational labeling. The human mind needs concepts and language in order to think and plan. Classifying persons according to significant characteristics and giving each group a name is essential to conceptualization and to planning effective educational treatments. Our problem in the past has not been that we have done too much labeling. Our problem has been that there have been too few labels and they have been too crude. We have grouped a large number of children with widely different characteristics and very different needs under one label, the mentally retarded, and have given them an undifferentiated program. What is needed is a more sensitive system for identifying children in need of specific education programs and a whole continuum of special education programs carefully targeted for children with specific needs.

Special education programs should be planned on the premise that every child is kept in the educational mainstream if at all possible. The self-contained classroom should be a treatment reserved only for the comprehensively retarded.

Figure 3.—Schema for Conceptualizing a Continuum of Special Education Services

Special educational program:
- Institutionalization.
- Self-contained special education school and program.
- Self-contained special education classroom and program.
- Split regular special education program: Music, art, P.E. etc. in regular class.

Regular academic program:
- Regular classroom with resource teachers:
  - Daily sessions with special resource teachers, crisis rooms, etc.
  - A few hours a week in resource rooms, remidial reading, bilingual program, etc.
- Regular classroom full time:
  - Tutors, cross-age teaching, programed learning, etc.
  - No special help.

Figure 3 presents, schematically, how such a continuum of special education programing might look. At the far right are those children in the regular classroom who need no special help beyond the regular classroom program. The next group, to the left, are those children who can be maintained full time in the regular classroom if they are given some additional individual help by tutors, mother helpers, cross-age tutors, or other persons working under the direction of the regular classroom teacher.

The next group consists of those children who need more intensive assistance with a special education teacher for a few hours a week outside of the regular classroom. Children in this group would be those who need remedial reading, English as a second language, or other

types of programs requiring special teaching skills. Closely related to this group are those children who may have regular, daily periods in a resource room, a crisis room, or other special program but are still, primarily, enrolled in the regular academic classroom.

The four categories to the extreme left are heavily weighted toward special education. However, even within the special education program, there can be differentiated treatments. Some children may have a program split between regular and special classes, sharing music, art, physical education, and other nonacademic classes with the regular students. The comprehensively retarded would spend their entire day in a self-contained special education classroom and program while still remaining in the same school building with their brothers and sisters and neighbors and friends. Only the comprehensively retarded with physical handicaps or other needs which require a specially designed physical plant would be isolated in self-contained special schools or institutions.

One of the most distressing current developments in special education in some regions of the United States has been the precipitous reassignment of many children to the regular classroom program and self-contained classrooms with no provisions for a continuum of special education services to meet their needs. It would be a great leap backwards if, as a result of modifying assessment procedures, we eliminate programs needed to serve children. Relabeling is not some magic panacea which suddenly enables children from socioculturally nonmodal environments to achieve in a regular educational program. It is essential that the children who received and were eligible for special education in the past continue to receive special education. It is essential that money be provided to continue support for special education programs for them. Changes in the types of assessment and the types of programing must not be used as excuses for saving money by eliminating programs. Instead, the financial support and the effort of special education teachers should be redirected into providing a wider variety of special services and programs geared to keeping as many children as possible in the educational mainstream and educating each child to his own maximum potential.

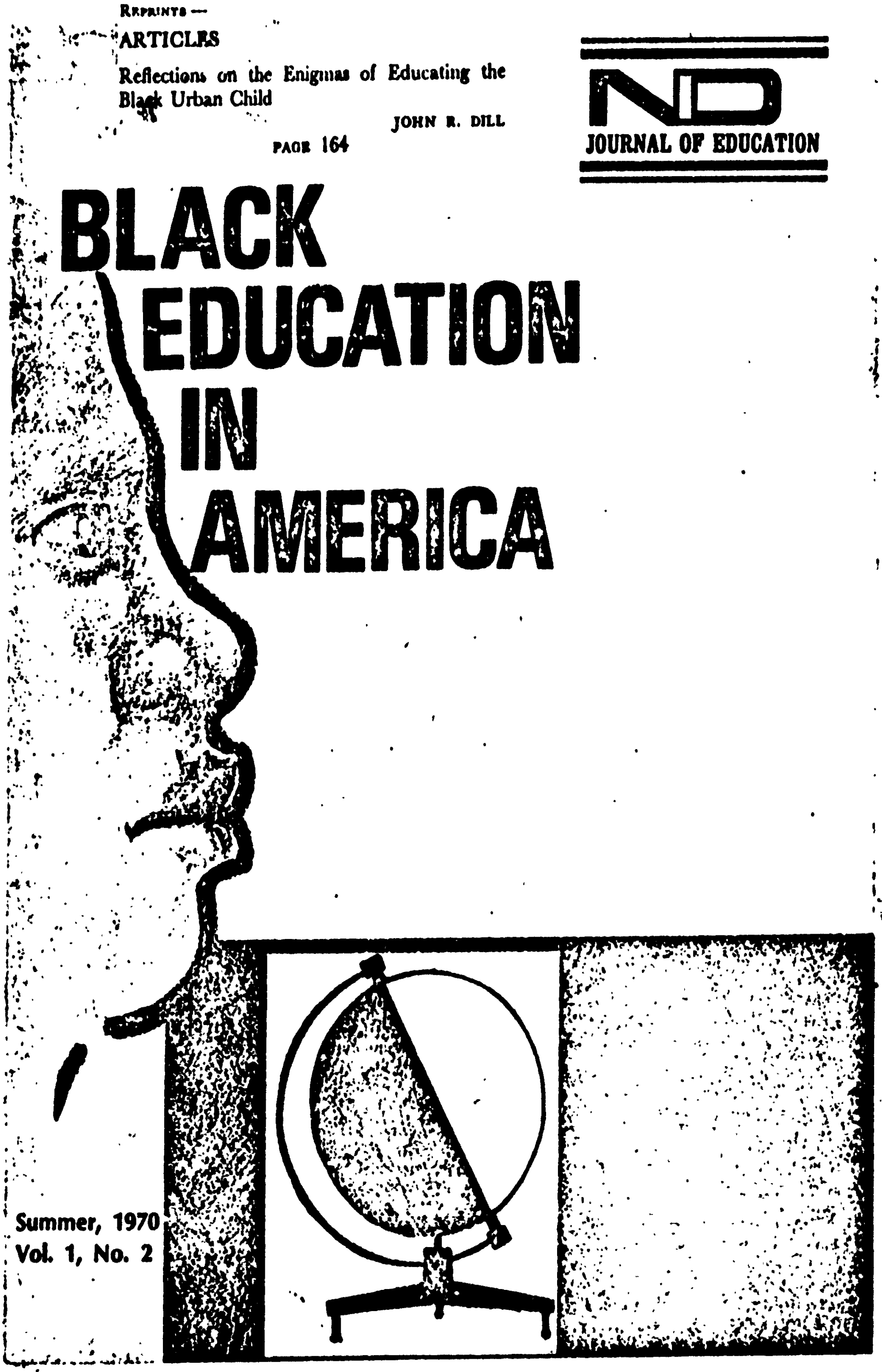
REFERENCES

Heber, R. F. A manual on terminology and classification in mental retardation. American Journal of Mental Deficiency, 1961, 64. Monograph Supplement. (2nd ed)

Terman, L.M., & Merrill, M.A. Stanford-Binet Intelligence Scale. Boston: Houghton Mifflin, 1960.

Wechsler, D. The measurement and appraisal of adult intelligence. (4th ed.) Baltimore: Williams & Wilkins, 1958.

Miscellaneous Publications



Reflections on the Enigmas of Educating the Black Urban Child

JOHN R. DILL
Research Scientist, Institute for Developmental Studies
New York University

The monumental task of providing meaningful and viable educational encounters for the Black child is becoming a major social problem in our urban centers. The traditional social and educational practices that were successful in the past are no longer operative; the learning processes that formerly characterized our educational system are faltering; even some innovative roles of the school as a social institution are exercises in futility.

As we approach the second decade of concentrated energy and interest in educating the urban Black child, it is appropriate to reflect on the chameleon-like nature of urban education with specific reference to the enrichment movement. Furthermore, it is crucial to present workable alternatives for future directions of urban educational reform. Thus the twofold purpose of this paper is to examine some of the global interdependent factors characterizing urban educational programs and to provide components of a futuristic model for educating the Black child. Although the major emphasis will be on the poor Black child who resides in the slums of our cities, many of the concepts in this article apply to the experiential conditions shared by children of other ethnic groups.

THE NOSOLOGY OF POVERTY

Through the national discovery of the needs of poor children, by such early writers as Riessman and Harrington, the advent of educational programming specifically geared to poverty childr was launched.[1] Federal and private sectors began to provide extensive funding for the express purpose of eradicating, reversing and preventing the debilitating effects of social, political and economic disadvantages on school performance.

Based on Riessman's classic, *The Culturally Deprived Child*, an unfortunate nosology evolved which hampered the course of educational advancement and attenuated its impact on the amelioration of social problems. This nosology resulted in an inappropriate and ill-fated alignment of the poor Black child's behaviors with those seen in the mentally retarded, emotionally disturbed and the physically handicapped. The assumption plainly was that the poor child's failure in school was due to factors similar to those of children suffering from the aforementioned maladies. Related to this phenomenon, one should realize that it was not fortuitous that many of the psychologists and educators, previously concentrating on the exceptional child (i.e., retarded or handicapped), began

[1] F. Riessman, *The Culturally Deprived Child*, New York: Harper & Row, 1962; M. Harrington, *The Other America*, New York: Macmillan, 1962.

to focus and shift their research interests to this very exciting area of research. After all, this was where the money was!

This host of symptoms contributed greatly to educational philosophy and social policy so that it was assumed and widely accepted that the Black child of poverty displayed in his behavior blatant deviancy or deficiency — not in an absolute sense — but always in reference to his middle-class peer. Although it would be futile to review the "symptoms" that characterized the poor child, suffice it to say that this gross catalogue was, and still is, a major hindrance in providing viability in the educational process for the urban Black child. Coupled with the value-ladened *in vitro* analyses of various minority cultures and their people (especially Blacks), educational programming took on a calamitous perspective.

It is revealing that only a few social scientists began to question the reality of social deprivation. Although the evidence based on comparative research only indicated a variable syndrome of experiential differences based on social class and race membership, there has been no conclusive and exacting empirical evidence that the impoverished child is deprived in a fashion similar to the classic sensory deprivation studies, as various writers seemed to espouse.[8] However, it was evident that educators were all too willing to embrace this premise. Secondly, it was almost totally forgotten that the repertory of behaviors displayed by this child is part and parcel of a realistic and intelligent pattern of survival skills in an impoverished environment. The middle-class norms of behavior would be completely inoperable in a poverty milieu. In addition, modern social scientists treated the Southern tradition of many urban Black emigrés with a cynical disdain never directed toward Western European immigrants.

The Black Family

In this regard, social scientists have attempted to search for the focal weakness of the scholastic failure of the Black child. While the school is becoming the most recent object of analysis, the perennial target of quasi-scientific study has been the Black family. E. Franklin Frazier, the noted Black sociologist, provided the impetus for subsequent writings that emphasized a pejorative viewpoint of the Black family structure.[9] Using Frazier's concept of familial disorganization among lower-class Blacks, one could interpret him as placing blame on the family for the child's failure in school. Even more invidious in its indictment deriving from a Frazierian thesis, was the infamous Moynihan Report proposing that:

> . . . At the heart of the deterioration of the fabric of Negro society is the deterioration of the Negro family. It is a fundamental source of weak-

[8] M. Deutsch, "The Disadvantaged Child and the Learning Process." In A. H. Passow, ed., *Education in Depressed Areas*, New York: Teachers College Press, 1963; M. Deutsch, "The Role of Social Class in Language Development and Cognition," *American Journal of Orthopsychiatry*, 35 (1965), 78-87; M. Deutsch, "Social Intervention and the Malleability of the Child." Paper presented at the Fourth Annual School of Education Lecture at Cornell University, Ithaca, New York, 1965.

[9] E. F. Frazier, *The Negro Family in Chicago*, Chicago: University of Chicago Press, 1932; E. F. Frazier, *The Negro Family in the United States* (rev. and enlarged ed.), Chicago: University of Chicago Press, 1966.

> ness of the Negro community at the present time. By contrast (to the white family), the family structure of the lower class Negro is highly unstable, and in many urban centers is approaching complete breakdown.[4]

No doubt Moynihan would have indicted the family for the child's academic failure, thus joining the legion of other social scientists who have done likewise. Parenthetically, it is enlightening that Andrew Billingsley has provided the focus of revising social science methodology so that more accurate analyses of the structure and functioning of the Black family can evolve.[5] Very few social scientists are able to take cognizance of the three invidious problems suggested by Billingsley that affect the Black family and the child in particular: poverty, prejudice and historical subjugation based on race.

From another perspective, the kind and quality of relevant stimulation the child receives at home during his first five years are still equivocal and open to precise research analysis. Although various researchers have collected data to indicate that the Black lower-class home has a scarcity of relevant materials (books, magazines, etc.) that could provide a wealth of stimulation for the child, it is still possible that a sparse social milieu can be distinctly enriching and provide the prerequisite intellectual skills essential to scholastic functioning.[6] In addition, an interactional analysis between Black parents and children as studied by Hess and Shipman has not directly determined the effect on subsequent school achievement, as many formulations would imply.

The Advent of Intervention

Thus, with a major focus on the denigrating circumstances surrounding the plight of the supposed disadvantaged milieu, interventive educational programs were launched. Some of the major assumptions of the intervention approach were: 1) the social environment is a major repository of knowledge and skills; 2) the "deprived" milieu has not provided the essential prerequisite experiences and skills necessary for scholastic success; 3) the earlier the intervention occurs, the better; and 4) any program is better than none.

Regardless of how sound and wise these assumptions of intervention, what evolved was a plethora of political machinations, institutional breakdowns and staff incompetencies, intertwined with the traditional complexities of sundry promises and undelivered practices. With only a casual survey of some of the major programs, one is able to detect a failure of allegiance to some basic theoretical formulations of social science.[7] For example, in those programs that

---

[4] D. P. Moynihan, *The Negro Family: The Case for National Action*, Washington, D. C.: United States Department of Labor, 1965, p. 5.

[5] A. Billingsley, *Black Families in White America*, Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

[6] M. Deutsch, *op. cit.*; R. Hess and V. Shipman, "Maternal Influences Upon Early Learning: The Cognitive Environments of Urban Pre-School Children." In R. Hess and R. Bear, eds., *Early Education*, Chicago: Aldine, 1968.

[7] C. B. Lavatelli, "Environmental Intervention in Infancy and Childhood." In M. Deutsch, I. Katz and A. Jensen, eds., *Social Class, Race, and Psychological Development*, New York: Holt, Rinehart and Winston, 1968.

trained mothers to adapt alternative methods of child rearing, the approach was doomed from the beginning. Considering the socio-historical-cultural matrix that to a large extent determines a person's behavior, it should not have been too surprising that mothers from second- or third-generation poverty backgrounds would be quite resistant to implementing any child-rearing techniques that were not endorsed and used by their mothers and grandmothers before them. Such training procedures are not simply the matter of "teaching an old dog new tricks" or other inappropriate euphemisms describing the retention of well-established old behaviors. Rather, one finds that the acceptance of new behaviors (as in this case, new child-rearing methods) can occur only after a thoroughly rational schema involving reinforcement and incorporating new behavior into a framework of existing behavior.

A cogent issue that interpenetrates the constructs of social intervention is: How pliable and modifiable is the child of poverty to the manipulation of his environmental experiences for the purposes of nurturing and stimulating his intellectual competence? Secondly, is the most effective antidote for the poor child fundamentally the provision of (or compensation for) those experiences that the child has missed? Thirdly, assuming the previous questions are answered affirmatively, is the school (as presently conceived) the appropriate social institution for providing such experiences? From an interdependent analysis of these queries, it is evident that educational planners were all too willing to affirm the first and second issues without serious consideration of the third. We must realize that the burden of providing remediative-intervention inputs to reverse the child's "deprivation" is greatly dependent upon formidable deliverance of services that the urban school has historically been unable to provide. One can also question whether or not the school structure as formulated for the near future must be radically restructured and redesigned to create a new concept of educational facilities.

Since the recent literature has fully documented the fallacies, incompetences and failures of the school system, one must acknowledge realistically the limits for the school's capability and responsibility in the educational framework. On a most basic level, one may identify some conditions that can occur in the school for urban children. Martin Deutsch, in a classic paper on "Social Intervention and Malleability," argues that the responsibility of an intervention effort is to

> diagnose the deficits (of the children) and attempt to determine the kinds of procedures that ameliorate them. Part of both the diagnosis and the determination might involve an evaluation of what is missing from the child's background, followed by a study of how the missing parts might have contributed to his more advanced development.[8]

It is unfortunate that such a basic yet highly crucial formulation has not been taken into serious consideration by educators. Of course such a proposal would require profound alterations in educational services that previously have been rejected in many social sectors.

---

[8] M. Deutsch, *op. cit.*, p. 11.

The Preschool Bandwagon

Based on the premise that the first four years in a child's life are crucial in terms of cognitive and affective growth, the energies and interests in educational circles were focussed toward the preschool period. Two important theoreticians, Benjamin Bloom (1964) and J. McV. Hunt (1961) have made significant contributions to this focus.[9] Bloom provoked interest by arguing that fifty per cent of the variance of intelligence (measured at age 17) can be accounted for by age four. Similarly, Hunt emphasized the significance of the early years when he eloquently described the importance of a "match" between the individual's behavioral competence and the environmental encounters.

It is no wonder that American educators were overly enthusiastic in their willingness to embrace the epistemological formulations of Jean Piaget, for it was seen that through Piaget the salvation of educational failure in the early childhood years was sought. However, the romance with the Genevan psychologist has been a confused infatuation, similar to other faddish tinkerings characterizing educational reform in the 1950's and 1960's.

Although a variety of programs appeared on the national scene, the only program that bore some degree of uniformity was the Head Start program, administered under the Office of Economic Opportunity and initiated in the summer of 1965. It should be emphasized that the Head Start program after its inception maintained a semblance of uniformity in name only. The lack of quality control erupted into giving Head Start a national image of ill-repute that was recently highlighted by the highly controversial national evaluation by the Westinghouse Corporation.[10]

Perhaps the most revealing comments on the nature of the Head Start movement are offered by Lavatelli who posits that this national program was

> designed, not to try out experimental approaches to the question of how to educate the disadvantaged, but to provide compensatory experiences based upon a traditional but enriched nursery school curriculum. . . . There appears to be general agreement that the traditional middle-class nursery school curriculum cannot bring the lower-class child up to the developmental level of the middle-class child.[11]

If Lavatelli's statements accurately express the goals and purposes of the Head Start program, one cannot help but challenge the reams of assessment evidence that are currently popular in evaluating programs. With major reliance on test data, measurable test score gains are used as evidence of the efficacy of a program. This unexpressed goal of boosting a child's performance level to that of the national norm, or conversely, the failure to do so, often unfortunately results in a pernicious misinterpretation of various programs as has been recently perpetrated by Jensen (1969) who asserts that "compensatory education has been

[9] B. Bloom, *Stability and Change in Human Characteristics*, New York: John Wiley, 1964; J. McV. Hunt, *Intelligence and Experience*, New York: Ronald, 1961.

[10] V. Cicirelli *et al.*, *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*, Washington, D.C.: Office of Educational Opportunity, 1969.

[11] C. B. Lavatelli, *op. cit.*, p. 359.

tried and it apparently has failed."[12] The only failure has been Jensen's in not realizing that compensatory education has never really been tried!

"Fakelores" About the Black Child

But education did not need Jensen to maintain the host of antediluvian and false concepts (i.e., "fakelores") about educating Black children. These fallacies, although not openly expressed, are shared by a great many educators who assert: "These children just cannot learn." A few of the most prominent "fakelores" are: homogeneity in Black children, the "unmodifiability" of the child, the notion of a fixed intelligence and the deficiency of Black language.

The concept of homogeneity in this case refers to the assumption that within a designated group of children, there exists a uniformity and similarity of attitudes, capacities, learning styles, etc., that typify the specific group of children. An extension of homogeneity is ability grouping which the *Dictionary of Education* defines as "the classification of pupils for the purpose of forming instructional groups having a relatively high degree of similarity in regard to certain factors that affect learning."

Although ability grouping based on homogeneity has been practiced since the turn of the century, it was principally geared to the effects on gifted students who could benefit from accelerated learning encounters. However, with lower-income Black children, homogeneity was contorted to assume that the impoverished child is accurately characterized by a host of behavioral symptoms that are almost universally displayed by all children from this milieu.[13] This distortion undertook a transformed benevolence by Riessman who presented the "overlooked positives" of the disadvantaged by presenting a number of so-called positive behaviors that at times could be taken as an apologetic "copout" for social institutional inertia.[14] For example, Riessman discusses the slow style of disadvantaged children as a strength. Yet, he fails to realize that this slowness is considered a weakness because the entrenched value system of the dominant society has little tolerance for this style.

Global descriptions of homogeneity either placed in a positive or negative context are devoid of any exacting prescriptions allowing or encouraging the individuation of a particular child. In fact, when one peruses the research literature, a major indication of the vacuity of homogeneity is evident. In reports on achievement and intelligence test data in a disadvantaged sample, mean test scores are generally lower than the average middle-class groups, but there is a tendency for a wide range of scoring within the disadvantaged group. These scores imply some heterogeneity of abilities within the disadvantaged population that at a very minimum should merit individualized instruction instead of the grossly inadequate teaching that significantly contributes to the high dropout

---

[12] A. Jensen, "How Much Can We Boost I. Q. and Scholastic Achievement?" *Harvard Educational Review*, 39 (1969), 2.

[13] R. Havighurst, "Who Are the Socially Disadvantaged?" *Journal of Negro Education*, XXXIII (1964), 210-217; M. Black, "Characteristics of the Culturally Disadvantaged Child," *The Reading Teacher*, 18 (1965), 465-470.

[14] F. Riessman, "The Overlooked Positives of Disadvantaged Groups," *Journal of Negro Education*, XXXIV (1965), 225-231.

rates of urban Black children on a psychological level in elementary school and physically by high school. Superficially, homogeneity can be a temporary panacea for the problem of individual differences if one buys the assumption that the differences among Black lower-class pupils (or any pupils) are reduced when there is some division into pupil groups based on a single test score — for example, reading achievement. In fact, taking reading as an example, one finds that there are such multiple skills as word and letter recognition, vocal responses, etc. — all clearly attesting to the superficiality of the ability grouping concept based on test score or other gross indices. More importantly, the concept of homogeneity as applied to the urban Black child, can imply that these children are uniformly deficient, immature or retarded, thereby allowing for little recognition of and programming for the extraordinarily high heterogeneity that one finds in any urban Black classroom. Although there have been meager attempts to incorporate this degree of high heterogeneity through individualized instruction, small groupings and extra teacher personnel, the needs have far outreached the demands. Thus, it becomes crucial in educating urban Black children to focus all innovative procedures that provide opportunity fo- the "match" between the child's behavioral competence and environmental encounters.

Related to the "fakelore" of homogeneity in the Black lower-class child is the notion that these children cannot benefit from instruction. Although such an indictment typically comes from either an immutable or inexperienced teacher, there does exist a general tendency to consider these children as academic failures. Much of this pernicious dialogue can be accounted for by the prejudices and expectancies that the educator brings to the school system. However to dismiss this as mere attitudinal predisposition could be quite superficial. Realistically, one can see that not only is this child's behavior modifiable but often the direction of its modification is antithetical to what the school wishes to endorse. The children might engage in activities that teachers consider as disruptive, they might be overly garrulous or their approaches to problem tasks might take on divergent (but often creative) directions.

The lower-class Black child, a victim of poverty, enters school with a different (not necessarily deficient) set of skills, capacities and motivations. The school culture, couched in a totalitarian insistence for submission to authority, cannot tolerate this child's style of behavior. It is not surprising then that Black children who display behaviors that "break the rule" of expected demeanor are often labelled as deviant, disruptive or mentally retarded. An illumination of hope, however, is provided by Herbert Kohl (1970) who advocates an open classroom as an alternative in combatting authoritarian education and preventing casualties.

One of the major obsessions in education has been intelligence and its measurement. The preoccupation with intelligence has a long history starting with Alfred Binet, the French psychologist, who is credited as the father of the intelligence test as we currently know it. American educators willingly endorsed or used Binet's products but failed to accept the bulk of his theoretical formulations. Such was the case when Binet warned against accepting the concept of a fixed intelligence where the individual's abilities were considered as inherent and

predetermined. It was obvious that American educators concentrated their efforts on the classification and placement of children rather than on the more crucial theories of developing intelligence. Interestingly, the cyclical polemics of intelligence have taken on new dimensions in the recent and controversial article by Arthur Jensen which supports a genetic determination to racial differences in intelligence.[18] It is highly unfortunate that an abundance of our time has been expended in circumlocutory and unconfirmed esoterica.

But how does this relate to educating the Black child? First, it has fallaciously been assumed in many educational circles that the Black child is limited by his inferiority in genetic predisposition. This postulate, although not openly expressed, is shared by many social scientists and seems to circumscribe much educational intervention. Any careful examination of the educational history of Black Americans in the 20th century would not support this contention of inherent inferiority. Whether it took the form of Booker T. Washington's industrial education at the cost of its more significant academic counterpart, or whether it was transformed into hastily constructed, superficially implemented compensatory programs of the last five years, the accurate assumption should be that Black children have been systematically denied intellectual stimulation because of the absence of any equitable or enriched educational stimulation. It is noteworthy to observe in this regard that the same casualness of urban educational programs geared to the poor and Black populations would not have been feasible in the nation's space program because our society would not have tolerated such maladies in our technological achievements as it has allowed in the educational sector. In addition, a major misorientation of American educational philosophy has been geared to the rarefication of human intelligence — defined and delineated by intelligence testing rather than an individual's actualization and contribution in society. This, in turn, has evolved into a voluminous research literature advocating Black intellectual inferiority. Thus, the emphasis on fixed intelligence and the concomitant advocacy of Black inferiority have led to a philosophy of educational constructs that nurtured institutional racism thereby impeding any major educational advancement for the Black child.

The language patterns and abilities that the Black child brings to the school context have been a controversial area of discourse among psychologists, linguists and educators. Much of the early research in the 1960's has focussed on social class and racial differences measuring such components as morphology, syntax, dialect and development. Although the results of these studies are far from being unidimensional, the findings generally are that Black children from lower socioeconomic status groups do "poorly" on various measures of linguistic competency. Parenthetically, it should be noted that these children's performances are uneven and dependent on the types of measures taken and the instruments used. Since language is intimately related to cognitive functioning, it is not surprising that it has been greatly emphasized in the newer curricula. However, one must continue to wonder if the supposed deficiencies of the lower-class Black child are really deficiencies or merely differences. And if they *are* deficiencies, does this presume a cognitive liability for the child? Probably not!

[18] A. Jensen, *op. cit.*, 1-123.

Some researchers have challenged the "deficiency" stance by advocating that the language of Black children represents a highly structured, meaningful, yet different language system.[16] The advocates of this position often suggest that teaching English as a second language is a major procedure to accelerate and stimulate language development (and eventually intelligence) in the Black child. Other theorists holding this position suggest the maintenance of the subcultural speech patterns and dialects thereby allowing some creditability for their existence. Such procedures must be considered primarily as racist in nature since it lessens the load of the school in particular, and society in general, in fulfilling its basic responsibilities of teaching and allowing children to enter the mainstream of society. As Deutsch comments on those advocates who would press for the maintenance of ethnic dialects:

> While a different dialect may be just as adequate or even superior for communication among its speakers as is the language of the broader culture, to accord the dialect the same position with respect to its economic or occupational role is to support an unreal and phony equalitarianism whose effect will be to keep the dialect-speaking subgroup in an inferior position within the broader society.[17]

The Future

What lies ahead for the educational enterprise in which Black children are concerned? Throughout the nation, Black communities are increasingly activating their energies and channeling them in the direction of configurational and operational change. The deplorable but unavoidable problem is that the school has become a political football between the communities and the establishment in a situation where the children have been the principal losers. Such skirmishes as personnel assignments, the power of community governing boards and the issue of a "Black" curriculum have increased the already existing tension in the school, but are necessary in order to develop a creditable education for Black children. One can expect, however, that cognizance of quality educational programming in the Black community will occur through an evolutionary (if not revolutionary) process whereby the substantive issues of the learning process will be reformed and implemented. It would be appropriate at this juncture to offer one *caveat emptor* to conscientious community groups: they must be careful in not accepting and practicing those highly tempting "faddish innovations" that have characterized the educational marketplace in the last few years. Such concepts as overstructured curricula, programmed materials and quasi-therapeutic methods geared to affective growth are by-products of radical educational reformists whose interests infrequently are for the benefit of the Black child and whose methods are more appropriate for the middle-class White child. One must

[16] W. Labov, "The Logic of Nonstandard Dialect." In J. Alatis, ed., *School of Languages and Linguistics Monograph Series, No. 22*, Georgetown University, 1969, 1-43; W. Steward, "Historical and Structural Bases for the Recognition of Negro Dialect," *ibid.*, 239-247; S. Baratz and J. Baratz, "Early Childhood Intervention: The Social Science Base of Institutional Racism," *Harvard Educational Review*, 40 (1970), 29-50.

[17] M. Deutsch, "Perspectives on the Education of the Urban Child." In A. H. Passow, ed., *Education in Depressed Areas, 2.* New York: Teachers College Press, 1970, p. 25.

not assume that this commentary is an indictment of the type of educational reform or experimentation that is sorely needed in our urban schools. However, it is the wise educator, parent or community activist who is aware of the lure and promises of such innovations that rarely result in significant change.

At this point, it is important to provide a general programming model for educating Black children. Centrally, the development of cognitive skills and strategies that are traditional in the school context is a priority. Using the Black model, emphasis should be placed on the nurturance of values/motives inherent to the learning process (i.e., learning for learning's sake).[18] From this juncture, there occurs an evolution of developing positive orientations to the specific cognitive skills and operations. What evolves is the enhancement of strategies, skills, creativity, etc., and other mechanisms that the school previously has failed to endorse. A basic component of educating Black children is improving their self-concept. This task could be addressed through direct learning experiences of the cultural heritage of Black people with paramount importance placed on Africa and other Black countries. Of course, it is realized that self-concept defined in ethnic dimensions is only one component of this diverse personality construct. These other components would be dealt with as they relate to the learning process.

When issues of staffing are concerned, priority is often given to the polemics about the relative effectiveness of Black and White teachers on Black children. Since no empirical evidence provides a clear-cut solution to this problem, it must be considered at this point that the teacher's ethnic membership may be relatively unimportant. Therefore, we must seek to identify additional indices of the effective teacher. However, one must challenge the traditional roles of adults in the classroom by advocating that personnel previously considered to be unequipped for teaching should be used in many roles. Quite plainly, schools must call on community persons (in addition to so-called paraprofessionals) to contribute to the learning process.

Commentators on education frequently allude to the need to meet the challenges of a complex technological society of tomorrow. Reference is usually made to the need for educational reform in order to insure all children egalitarian participatory rights in such a cultural matrix. What they fail to mention is that this societal transition is neither in the near nor remote future but is presently among us. This all points to the fact that once again, the poor, and particularly the Black child, who is presently suffering from educational bankruptcy, could also be afflicted in the near and remote future. As Black communities are demanding a quality education that society considers as a basic right to which all children are entitled (e.g., reading on grade-level skills expected at developmental levels, etc.), the educational establishment must begin to analyze its past failures and rechannel its future directions.

Related to these demands is the need to develop an alliance of those concerned in a revolutionary educational enterprise that is characterized by vitality and achievement. Perhaps, a priority that should be considered in fostering

[18] L. Black, "A Position Paper Concerning Our Projected Learning Experiences." Paper presented at the Black Child Development Conference, Washington, D.C., June, 1970.

significant educational experiences for the Black child is a powerful coalition of Black professionals who share a primary interest in the educational system as it affects Black children. Banks considers such a coalition of Black social scientists whose organized efforts would

> search for better methods of teaching reading other than contribute to the growing list of publications with pictures of Black children and stories of city life; create and standardize appropriate test devices so that we can tap the talents of Black children rather than join those who advocate the elimination of tests; and develop academically oriented preschools to give our children a real head start in terms of skills rather than allow such programs to serve in a hit or miss fashion and provide data that demonstrate their ineffectiveness.[19]

To achieve such objectives of course requires a complete overhaul and upheaval of our social sciences as they presently exist: programs cannot be merely and haphazardly attempted and then dropped; research methodology which is more utilitarian in a laboratory setting must be replaced by more precise procedure or totally abandoned; the marriage between teacher and researcher must be commenced; and parents and the school establishment must initiate a dialogue in order to implement a "mutual assimilation" into each other's social-psychological-cultural milieu.

---

[19] R. Banks, "New Directions in Education — 1970," *Freedomways*, 10 (1970), 14.

77-942 O - 72 - 30

[From New Society, June 17, 1971]

RACE, INTELLIGENCE, AND EDUCATION

By H. J. Eysenck, Professor of Psychology, Maudsley Hospital, London

Debate continues over Jensen's paper on heredity and intelligence. But now, with the heat dying down, may be a good time to question some of the counter-arguments. The concepts in the title of this article tend to arouse an unusual degree of emotion, as the reception of Arthur Jensen's [1] well-known monograph "Environment, heredity and intelligence," in the *Harvard Educational Review*, demonstrated only too well. But I would like to try to dispel some of the erroneous notions which have accumulated about this subject, and to review briefly and objectively the evidence which various groups of scientists have gathered over the years.

This aim is altogether different from that which inspired Jensen; in his monograph, the racial problem occupied only a very minor place, and he never attempted to review the literature concerning it in any detail. Audrey Shuey's [2] book, *The Testing of Negro Intelligence*, is really much more relevant—but, though it was published in 1966, it has not been widely read in this country. She, too, however, only dealt with one part of the evidence—namely that concerned with environmental factors (Jensen was much more concerned with the genetic evidence).

Many people may feel that even raising the possibility of the genetic determination of American Negro inferiority in IQ tests—let alone suggesting that it might have to be answered in the affirmative—is a form of racism, and that anyone holding such views would automatically be in favour of segregation and other forms of discrimination. But it is important to realise that the *scientific problem* of the degree to which various groups are inferior (or superior) to one another on IQ tests, and the determination of the causes of such inferiority or superiority, is quite distinct from the *social and ethical problems* which confront us.

Even if it were a fact that American Negroes are genetically predisposed to score less highly than whites or Orientals on such tests, it might still be argued that socially and ethically the right thing to do was to give special care, consideration and enriched education to Negro children, in order to try, as far as possible, to bring them up to the average. Indeed, anyone aware of the terrible things that white men have done to coloured ones, over the centuries, would find it diffi-

[1] A. R. Jensen, "Environment heredity and intelligence" (*Harvard Educational Review*, vol. 39, 1969).

[2] A. M. Shuey, *The Testing of Negro Intelligence* (Social Science Press, 1966).

cult to deny the justice of the black man's claims to restitution. But however we answer this second question (and my own answer, like Jensen's and Shuey's, would be emphatically against segregation and in favour of reparation), it is not the same question as the first; and answering the factual question should not be influenced by social and ethical considerations appropriate to the second.

Many researchers argue that the very notion of "race" is unscientific, and that we should rather be talking about ethnic subgroups. Certainly, it is erroneous to think of races as immutable types, qualitatively differentiated from each other. It is more correct to think of them as groups or populations quantitatively distinguished along a number of different dimensions. The discovery and elaboration of serological genetics, by means of which specific genes can be recognised by chemical reactions with the components of human blood, and the constant finding of new blood-genes, has forced experts to expand the number of ethnic subgroups very greatly. Different gene-pools can be found even within relatively small countries. (Thus, there are many gene-pools in Wales alone.)

Even apparently homogeneous groups, such as the American Negroes, can be shown to be quite heterogeneous in many ways. North American Negroes, to take but one example, are much more hybridised than South American Negroes, and it is even possible to assess the degree of hybridisation quantitatively. Similarly, American Negroes are not identical with African Negroes, and these in turn constitute many different subgroups. Nevertheless, it is not meaningless to ask whether American Negroes, on the average, score less highly on IQ tests than American whites do, or American Orientals, or American Indians. These relatively large groups, do show quite pronounced serological and morphological differences from each other, which swamp within-group differences.

However, we must be careful not to use American Negroes as a prototype for all Negroes, or American whites as the prototype for all whites. American Negroes score, on average, some 15 IQ points below American whites, but so do some white groups outside America. The Irish, for instance, are at about the same level as the American Negroes. (The cause, in the case of the Irish, is probably the constant braindrain through emigration, and the holocaust of their brightest and bravest men inflicted upon them by centuries of English aggression.) Similarly, there is the possibility that the low scores of American Negroes on IQ tests may be due to selective immigration (the duller Negroes being captured, or sold, to the slavers) and to the killing-off of "uppity" bright Negroes who didn't suit their masters' desire for dull beasts of burden.

How can we account for the fact that, for some 60 years now, psychologists who have tested American Negroes, have found them to average about 85 IQ points, compared with the white population's 100 points? Southern Negroes score somewhat lower even than this, northern Negroes somewhat higher, but, on the whole, the figure has remained remarkably steady.

Environmentalists maintain that this is due entirely to the poorer conditions under which Negroes are brought up and have to live: poor

food, poor parental care, poor education, poor jobs, unemployment—the list is endless. They often add such factors as black children responding poorly to white testers; blacks doing poorly on white testers' tests, because these are biased in favour of white culture generally; and poor motivation of black children due to pervasive experiences of failure.

By contrast, geneticists maintain that genetic factors are likely to play a part in addition to any environmental ones. Geneticists, these days, never maintain that genetic factors are all-important. Such a belief would go directly counter to the fundamental distinction geneticists make between "genotype" and "phenotype." The genetic hypothesis therefore admits, without argument, that part of the difference observed is likely to be due to environmental factors. It is very important to bear this in mind in appreciating the argument that follows.

It is possible to test the value of each protagonist's arguments by arranging experiments which will tell us to what extent hypothesis is in accordance with fact. Jensen concentrated on deductions from the geneticist's theory; but this is probably the weakest way to support that hypothesis. Demonstrations that individual differences in intelligence are very strongly determined by heredity—while correct and important—apply only within a given racial group; they do not necessarily carry over to racial differences.

There are weaknesses in the contrary argument, too; to suggest that what is true of white intelligence is not true of black intelligence, suggests even greater differences between white and black than the purely quantitative ones suggested by Jensen. But as many geneticists who have considered Jensen's attempt have pointed out, the evidence at the moment is suggestive but not conclusive. It is not impossible to produce conclusive evidence, but that is in the future. At the moment, one can only say that the genetic evidence is in good agreement with the genetic hypothesis, but is not compelling.

There is one interesting experiment which gives a kind of direct proof of racial differences in IQ based on genetic determinants, and which could be extended to furnish such proof on a larger scale, and in relation to other racial groups. M. M. de Lemos[3] tested groups of aboriginal children in Australia. These children lived in mission compounds, and were all educated in common, and treated alike. Search of the archives showed that some of these children had one white great-grandparent, unbeknownst to the children. When the children were tested on a Piaget-type of intelligence test, the children with part-white ancestry obtained very significantly higher scores than the others did. In America, too, there has been a tendency for Negroes with lighter skins (and therefore probably a greater admixture of white ancestry) to do better on IQ tests. But it is possible that environmental conditions are somewhat more favourable for lighter-skinned Negroes, so that not too much should be made of these figures.

It turns out to be more useful to test the deductions from the environmentalist theory, if only because these tests can be arranged

[3] M. M. de Lemos, "Development of conservation in aboriginal children" (*International Journal of Psychology*, No. 4, 1969)

fairly easily, and without having to undertake impossible breeding experiments.

Thus, it is easy to test groups of Negro children using white and black testers, and show that the colour of the tester is quite irrelevant; sometimes there is a slight advantage to the one, sometimes to the other.

Poor motivation, too, can be investigated. Jensen carried out one particularly interesting experiment in which two sets of tests were incorporated within a single format. To the testee the two tests looked exactly alike. But while one was a simple test of cooperation and motivation, the other was a genuine IQ test. Negro children did as well on the motivational test as white children did, but they did very much less well on the IQ test.

There is much other evidence to show that motivational factors do not play an important part in the results so far reported, and that Negroes have as high levels of aspiration as do whites. Thus empirical studies fail to support these quite specific environmentalist hypotheses.

When we turn to general socioeconomic status, we find other facts which are in disagreement with the environmentalistic viewpoint. We can compare American Negro and white children attending the same schools, and coming from identical environments. When we do this, the differences in IQ are reduced a little, but not very much. In fact, when middle class Negro children, going to good schools in American middle class neighbourhoods, are compared with working class white children, going to rather poor schools, then it is found that, in spite of their inferiority on all aspects of socio-economic status, the white children are still marginally ahead.

Such findings are difficult to explain on a purely environmentalistic hypothesis. It is equally difficult to explain the fact that American Indians, although as far below Negroes in respect to socioeconomic status as Negroes are below whites, are yet not inferior to them on IQ tests. If socioeconomic status is so important in producing IQ differences, why does it suddenly cease to be a factor when we carry out this comparison?

Oriental children in California come from families who are distinctly inferior to white families in socioeconomic status, but they do better on IQ tests than white children. How is this possible on an environmentalistic theory? The relationship between IQ and socioeconomic status is complex, and by no means as linear and straightforward as environmentalists suggest. Such simple theories completely fail to account for the majority of the observed facts.

We must now turn to another argument—which says, in effect, that tests of IQ are not measures of intelligence at all, but simply educational tests of cultural knowledge; that American Negroes do poorly on verbal tests and tests of knowledge because of their inferior education; and that if "culture-free" tests of intelligence could be devised, then Negroes would do as well as white, not being held back by environmental handicaps.

Alas, this argument, too, can be shown to be hopelessly wide of the truth. It is true that there are no culture-free tests, but tests differ in the degree to which they draw upon cultural knowledge, acquired through education. Typical tests of school knowledge contrast with

such "culture-fair" tests as Raven's Matrices. In these the testee is not required to draw upon any school-taught knowledge, but simply to educe relations and correlates when he is confronted with simple drawings of a nonrepresentational kind.

Within white groups, heredity seems to contribute only about a third as much to individual differences on the less culture-fair tests as compared with the Raven's Matrices type of test. We, thus, have tests which differ in their dependence on cultural knowledge. The environmentalist hypothesis would require American Negroes to be less inferior to whites, the more culture-fair the test happened to be.

Exactly the opposite is true. Negroes do best—compared to whites—on tests heavily contaminated with school knowledge. They do poorest on "culture-fair" types of pure intelligence-test items, involving the ability to think abstractly, and not calling for school knowledge in any form. The often-repeated notion that Negroes are inferior only on verbal types of items and tests is simply wrong. It is on these items that they do relatively well. It is the non-verbal, highly abstract types of test where American Negroes perform worst.

Curiously enough, the pattern of "good on abstract tests—poor on cultural and educational tests" fits the American-Mexican group very well. For them, it might be held that environmental disadvantages are responsbile for much of their poor showing. These children are well below the Negroes in socioeconomic status, but they test on the average about the same. (Why, on environmentalistic grounds, are they not inferior?) Yet they score *above* the Negroes on abstract "culture-fair" tests; and *below* them on cultural tests involving school knowledge. (American-Orientals, as I have already said, are inferior in socioeconomic status, but superior on abstract tests of IQ to all other groups. They are not superior to whites on culture-bound tests.)

Another argument that goes against the environmentalist hypothesis is that, if it was true that inferior schooling holds back the black child, then his inferiority on IQ tests should *increase* as he goes through school. Yet this is not so. The observed differences neither increase nor decrease, on the whole.

Why, despite all this evidence, are the environmentalist hypotheses so firmly held, defended with such tenacity, and given a charmed life where no experimentally ascertained facts are allowed to impinge on their sacred image? If we want to help the blacks achieve equality of *opportunity* (their entitlement to which I consider to be axiomatic), then we can do so only on a basis of factual knowledge. Pretence, however well-intentioned, must lead to failure and further frustration. The acknowledged failure of Headstart and other similar compensatory education programmes in the United States is a good example of hopes raised, only to be dashed to the ground.

This, I think, must be the answer to all those who prefer not to experiment in this field, for fear of what they might discover. If it is true that American Negroes are genetically predisposed to lower achievement on IQ tests—and there is no doubt that the evidence is, *prima facie*, quite strong on this point, even though in the nature of things it cannot be absolutely conclusive—then, it is not doing blacks any favour at all to pretend otherwise. At the moment, many American universities are engaged in what might be called inverse racist policies

of admission. Negroes are frequently admitted with qualifications very much below the minimum requirements for white students. This leads to their failure on orthodox courses, and the demand for special "black studies," or more lenient marking systems, or both.

It is wrong to refuse blacks admission on the grounds of colour. It seems to me equally objectionable to grant blacks admission on grounds of colour alone—i.e., without the usual intellectual and educational qualifications. If the one is racism, so is the other. Both treat blacks and whites as different groups. But the egalitarian doctrine asks us—rightly, in my submission—to regard each person as an individual, regardless of colour, sex, or religion.

Nor do such policies produce a lowering of racial barriers. Within the universities that practise this "inverse discrimination," the blacks tend to group themselves together, in opposition to the whites. This produces an exaggeration, rather than an amelioration, of the racist conflict. Easy, politically simple, policies like Headstart, and the lowering of entrance requirements for blacks, do not result in what we all want to see.

By refusing to face the facts the black's alleged friends may turn out to be his worst enemies. Only factual research, carried out with the aim of helping the Negro achieve his rightful aspirations, can give us the answer to the problems which at the moment seem insuperable. While we may not know anything like as much about these matters as we would like to know, it would be wrong to pretend that we know less than we do.

[From Phi Delta Kappan, January 1972]

I.Q. HERITABILITY, RACE DIFFERENCES, AND EDUCATIONAL RESEARCH

By N. L. Gage

N. L. Gage (2448, Stanford University Chapter) is professor of education and psychology at Stanford University and chairman of the Executive Board, Stanford Center for Research and Development in Teaching. He edited the classic Handbook of Research on Teaching for the American Educational Research Association and was AERA president in 1963-64. His Teacher Effectiveness and Teacher Education: The Search for a Scientific Basis has just been published by Pacific Books, Palo Alto, Calif. Mr. Gage says, "For quick help in sending literature to me at Stanford in Germany, where I wrote this paper, I am very grateful to Arthur Jensen, Henry Kaiser, and Douglas Pidgeon. To Mr. Pidgeon I am also grateful for suggestions of ways to clarify some sentences and for reminding me of Bloom's work, to which I refer."

The most important way to disprove Jensen's hypothesis is to reduce race differences through education and other kinds of environmental influence. Unless educators and others who influence the environment are effective in making Negroes as successful educationally and occupationally as whites, Jensen will for practical purposes have won the argument. The data are equivocal, as endless clashes of opinion among well-intentioned scientists indicate. But the opinions of scientists are not, in any event, our primary concern, whether we are black Americans or white. What we really want for our society is the elimination of racial differences in school and job success.

Jensen hypothesized that "genetic factors are strongly implicated in the average Negro-white intelligence difference * * *"[1] This hypothesis provokes intellectual and political fights because of its implications, if accepted, for legislation aimed at improving the education of Negroes and other low-income students. It can be construed as implying that such legislation is futile. If that belief is accepted by political leaders, our governments will not give educators, and their research and development arms, the money and other resources they need for work toward reducing the educational and employment disadvantages suffered by Negroes and other minorities. It is because of what Professors Jensen and Shockley say to the President and the Congress that educators, who want a fair chance to try their approach, should be concerned.

[1] Arthur R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?," *Harvard Educational Review*, Winter, 1969, pp. 1-123.

In what follows, I intend, first, to examine some underemphasized aspects of the data on whites that are cited in the controversy. Then I shall consider the relevance of these data, however valid they may be for North European and American whites, to the question of genetic determination of Negro-white differences in mean IQ. Third, I shall call attention to the need for the educational research and development that can produce necessary improvement in the achievement and attitudes of Negro youth and, ultimately, provide the only definitive test of Jensen's hypothesis.

IDENTICAL TWINS REARED APART

The logic of the problem of estimating genetic and environmental variance in IQ leads us to identical twins reared apart as the source of the most relevant data. Such twins provide natural experiments in which the genetic source of variance is eliminated. Hence any remaining differences in IQ between such twins can be attributed (apart from errors of measurement) to variance in environmental factors—ranging from the intrauterine position and nutrition of the fetus to the quality of the home, neighborhood, and school in which the child receives his education. It is little wonder that—although they also refer to data on identical twins reared together, and on fraternal twins and ordinary siblings reared together and apart, and on real and foster parents and children, and persons in other blood and environmental relationships—students of the problem find their most convincing data to be the IQs of identical twins reared apart. Thus, Jensen, subsequent to the storm aroused by his 1969 paper, published a new analysis of the four largest studies of such twins.[2]

But it is not enough merely to know that identical twins were reared apart. How "apart," or different, were their environments? If two identical twins were separated during their first 6 months of life and raised in different families, their environments may nevertheless have been very similar. If both were raised in the families of, say, college professors, our traditions would lead us to expect that both twins had "good" homes and received much intellectual stimulation. If both twins were raised in lower middle class homes, where both sets of foster parents had a high school education, we should again infer that these twins had fairly similar environments. If both twins were raised in isolated rural or urban slums by poor and uneducated people, without the books and conversation that we expect to foster intellectual development, then also they were not reared very "apart," even though they lived in houses separated by 3,000 miles.

These ideas about the meaning of "apart" are not new, but they have been underemphasized. Jensen devoted only 11 lines to the matter,[3] in describing the "most interesting" of the available studies (Cyril Burt's):

". . . most important, the separated twins were spread over the entire range of socioeconomic levels (based on classification in terms of the six socioeconomic categories of the English census), and there was a slight, though nonsignificant, negative correlation between the en-

---

[2] Arthur R. Jensen, "IQ's of Identical Twins Reared Apart," *Behavioral Genetics*, 1970, pp. 133–46.

[3] Jensen, *op. cit.*, 1969.

vironmental ratings of the separated twin pairs. When the twin pairs were rated for differences in the cultural conditions of their rearing, these differences correlated .26 with the differences in their IQs Differences in the material conditions of their homes correlated .16 with IQ differences. (The corresponding correlations for a measure of scholastic attainments were .74 and .37, respectively. The correlation between the twins in scholastic attainments was only .62, indicating a much lower heritability than for IQ.)" [4]

If we turn to Jensen's source,[5] we find that Jensen has chosen the lower of two correlations, i.e., the one for the individual intelligence test. The *r* for the group test is .43. But only a little more information becomes available in this source, despite the fact that Burt considered such correlations to be "the only satisfactory method" (p. 149) to demonstrate the importance of environmental opportunities. In his similarly slight treatment (12 lines) of these data, Burt wrote:

"For this purpose we have assessed the economic and cultural conditions of the homes in terms of a conventional scale similar to that employed for assessing intelligence and educational attainments, namely, one in which the mean is 100 and the standard deviation is 15. The correlations thus obtained are shown in Table 3 [our Table 1].

"It will be seen that differences in educational attainments are highly correlated with differences in cultural background. There is also a significant correlation between cultural differences and differences in the scores for the group test taken as they stand. But the correlations for the individual test and the final assessment are so low as to be nonsignificant with a sample of this size. The differences in educational attainments show a small but significant correlation with differences in material conditions, chiefly no doubt because the latter are responsible for differences in the children's physical health and school attendance." [6]

TABLE 1.—CORRELATIONS BETWEEN DIFFERENCES IN IQ AND DIFFERENCES IN HOME CONDITIONS FOR IDENTICAL TWINS REARED APART

[N-53 pairs]

Test results	Home conditions	
	Differences in cultural conditions	Differences in material conditions
Differences in intelligence:		
Group test	0.43	0.21
Individual test	.26	.16
Final assessment	.15	.18
School attainments	.74	.37

Note: Correlations over .29 are significantly different from zero ($P<.05$).

Source: Cyril Burt, "The Genetic Determination of Differences in Intelligence: A Study of Homozygotic Twins Reared Together and Apart," British Journal of Psychology, 1966, p. 149.

Several comments are warranted by this treatment of these data by Jensen and Burt, who was also a strong hereditarian. First, the treatments seem brief, almost grudging, in view of the admitted impor-

[4] *Ibid.*, p. 52.
[5] Cyril Burt, "The Genetic Determination of Differences in Intelligence: A Study Of Monozygotic Twins Reared Together and Apart," *British Journal of Psychology*, 1966, pp. 137–53.
[6] Burt, *op. cit.*, p. 149.

tance of the data and the pages devoted by these authors to other matters. No standard deviations or means of the environmental difference variable were provided; thus it is impossible to compute a regression equation for predicting differences in IQ from differences in environment for identical twins reared apart. Second, no distributions or lists of environmental difference data were given, although Jensen for his *Behavioral Genetics* article apparently obtained Burt's original IQ data. He provided lists of the IQs and various statistics based on them (mean, standard deviation, mean absolute difference between twins, standard deviation of the absolute difference, the intraclass correlation between twins, and the difference correlation between twins, which "indicates the degree of similarity between twins relative to the similarity between persons paired at random from the general population").[7]

Third, Burt qualified the correlation of .43 between differences in group test IQs and differences in cultural conditions by referring to lower correlations (.26 and .15) for individual test IQs and his so-called "final assessment" IQ, respectively. His suggestion seemed to be that the latter measures and the *r*s involving them are more valid than those based on group tests. But he gave no evidence to support such higher validity for the individual test and, despite much lore to the contrary, individual tests cannot be assumed without explicit evidence to be more valid.

The "final assessment" was obtained after the group and individual tests had been given and their results had been "submitted to the teachers [of the children] for comment or criticism; and wherever any question arose, the child was reexamined." [8] Since Burt defined intelligence as "innate general cognitive ability," it is not inconceivable that teachers were unintentionally influenced by him to criticize and question IQs out of line with children's hereditary backgrounds rather than those discrepant with the cultural conditions in their homes. Thus the "final assessment" could readily, even if unintentionally, have been biased in such a way as to reduce its tendency to reflect a child's environment and increase its conformity to the child's hereditary background. Without better information as to the nature of the "final assessment," we cannot appraise its validity. Hence its seems questionable to intrude the IQ differences based on "final assessment" into the consideration of relationships between such differences and cultural environment differences.

Fourth, we should take care not to regard the school attainment data in Table 1 as irrelevant to the social policy issues to which IQ data apply. For it is school (and occupational) attainment with which society, parents, and educators are most directly concerned. IQs are important only insofar as they predict school and occupational attainment. We are concerned that blacks have average IQs substantially below the average IQs of whites not because IQs are important in their own right but only because they throw light on ability to succeed in schools and jobs.

If environmental conditions of disadvantaged children can be arranged so as to improve school attainment, then such conditions are important regardless of their effect on IQ. They provide a basis for

---

[7] Jensen, *op. cit.*, 1970, p. 140.
[8] Burt, *op. cit.*, p. 140.

hoping that the educational and occupational discrepancies between blacks and whites can be reduced and then eliminated. Hence we stress the great importance of the correlation of .74 between differences in school attainments and differences in cultural conditions in the homes of the separately reared identical twins. This *r* suggests that, even when heredity is held constant, substantial differences in school attainment can be produced by environmental differences. Along with the correlation of .43 for the group-test IQ differences, Burt's results indicate that, given environments different or "apart" enough, we can produce major differences in IQ or, even more important, school attainment, even among persons with the same genetic composition.

Unfortunately, we have no units for measuring environments that tell exactly what exists or happens in them. So we cannot tell from Burt's data (or those of a similar study discussed below) just how much environmental difference of what kind was associated with a given difference in the IQs of the identical twins. This neglect of technical detail in reporting on the environment measures can perhaps be remedied by further examination of Burt's legacy of data. If so, we might have a basis for determining whether the environmental differences that produced 10-point or greater IQ differences in 14 of Burt's 53 identical twins [9] resemble those that have produced such differences between American blacks and whites.

Let us turn now to the only other study, by Newman, Freeman, and Holzinger in 1937,[10] in which IQ differences and environmental differences were studied to reveal correlations for identical twins reared apart. This study yielded the data shown in table 2. The scatterplot for the relationship between IQ differences and rated differences in educational advantage is shown in figure 1. As the note for table 2 indicates, the environmental-difference ratings were obtained from five judges who read the case material on each pair of twins.

It is clear that the IQ differences were larger for those twins whose estimated educational disadvantages differed more; the IQ difference and the educational-advantage difference correlate .79. The correlations between IQ difference and estimated social-advantage and physical-advantage differences were, respectively, .51 and .30. The fact that both the *r*s were lower than .79 suggests the special relevance of the education-advantage ratings. When considered along with the results of Burt's study, shown in table 1, where the correlations between IQ difference and "material-condition" difference were also smaller, the results indicate convincingly the association between differences in educational environment and differences in IQ even when genetic differences are held constant at zero.

Bloom carried the analysis of the data in table 2 a step further:

"We have divided the separated twins into two groups. For one group of 11 pairs, each pair of separated twins had very similar educational environments. The rank correlation for their IQ scores was +.91, whereas for the eight pairs that had the least similar educational environments, the rank correlation for their IQ scores was only +.24." [11]

---

[9] Jensen, *op. cit.* 1970, p. 14.

[10] H. H. Newman, F. N. Freeman, and K. J. Holzinger, *Twins: A Study of Heredity and Environment.* Chicago: University of Chicago Press, 1937.

[11] Benjamin S. Bloom, *Stability and Change in Human Characteristics,* New York: Wiley, 1964, p. 70n.

TABLE 2.—SOME DATA FROM IDENTICAL TWINS REARED APART

[Newman, Freeman, Holzinger, 1937; Muller, 1925; Gardner & Newman, 1940; Saudek, 1934] [1]

Case number	Sex	Age at—		Environmental differences			IQ difference
		Separation	Testing	1. In years of schooling	2. In estimated education advantages	3. In estimated social advantages	
11	Female	18 months	35	14	37	25	24
2	do	do	27	15	32	14	12
18	Male	1 year	27	4	28	31	19
4	Female	5 months	29	4	22	15	17
12	do	18 months	29	5	19	13	7
1	do	do	19	1	15	27	12
17	Male	2 years	14	0	15	15	10
8	Female	3 months	15	1	14	32	15
3	Male	2 months	23	1	12	15	—2
14	Female	6 months	39	0	12	15	—1
5	do	14 months	38	1	11	26	4
13	Male	1 month	19	0	11	16	1
10	Female	1 year	12	1	10	15	5
15	Male	do	26	2	9	7	1
7	do	1 month	13	0	9	27	—1
19	Female	6 years	41	0	9	14	—9
16	do	2 years	11	0	8	12	2
6	do	3 years	59	0	7	10	8
9	Male	1 month	19	0	7	14	9
Muller	Female	do	30	9	?	?	—1
Gardner & Newman	do	do	19	0	2	?	—3
Saudek	Male	do	20	0	?	?	+4

[1] From R. S. Woodworth, "Heredity and Environment: A Critical Survey." New York: Social Science Research Council, 1941.

Note: The estimated differences in educational and social advantages are in points, with a maximum possible of 50. From the case material each of five judges rated the envi onmental differences between every pair of twins on a scale of 10 points, and the figure given in the table is the sum of these five ratings. A minus sign before an IQ difference means that the twin received the higher rating for educational advantages obtained the lower IQ.

Bloom's analysis explicates what is implicit in the data presented in tables 1 and 2: The very high correlations, averaging .85,[12] between the IQs of identical twins reared apart result from the similarity of the environments in which the twins were reared. As the similarity becomes lower, the correlation becomes slower.

Jensen has shown that the absolute differences in IQ of identical twins reared apart fall into a distribution that "closely approximates the chi distribution,"[13] which would result from taking the absolute difference between pairs of values drawn at random from a normal distribution. This means in turn that environmental effects on the IQ as represented by cotwin differences, are normally distributed. Since all of the 122 pairs of identical twins reared apart that were involved in the four studies reviewed by Jensen were drawn from English, Danish, and North American Caucasian populations, this finding seems reasonable. At least, there are no obvious historical or sociological factors that would interfere with the operation of the large number of independent, relatively equally influential factors that make any set of data, including these environmental differences, fall into a normal distribution.

But such a normal distribution is not tantamount to a finding that the IQ differences are merely "chance" or "uncaused," attributable entirely to errors of measurement. For table 1 and figure 1 show that

[12] Jensen, *op. cit.*, 1970, p. 137.
[13] *Ibid.*, p. 141.

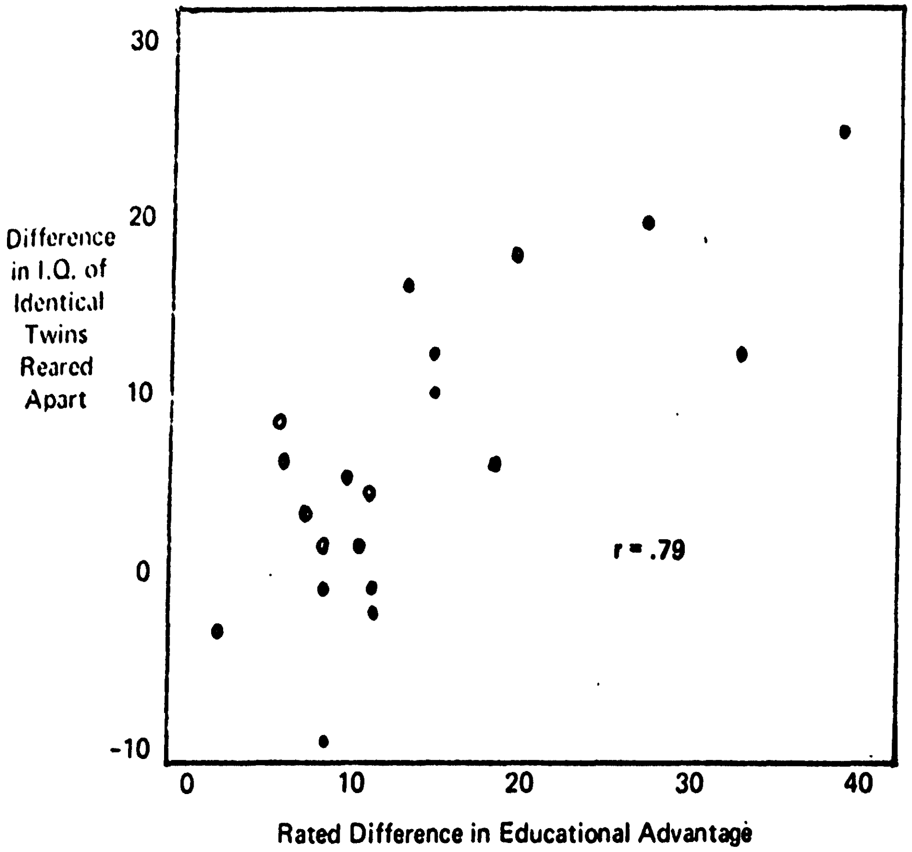


Figure 1.—Scatterplot Showing Relationship Between Differences in IQ and Educational Advantage of 20 Identical Twins Reared Apart. (Based on H. H. *Newman*, F. N. Freeman, and K. J. Holzinger, *Twins: A Study of Heredity and Environment*. Chicago: University of Chicago Press, 1937.)

the IQ differences are obviously associated with environment differences. Thus it is difficult to see the point of Jensen's citing the normality of distribution in relation to the frequently cited case of Gladys (IQ 92) and Helen (IQ 116) in the study by Newman et al. (This case is represented in figure 1 by the dot in the upper right-hand corner.) These twins "had markedly different health histories as children. . . . Gladys did not go beyond third grade in school, while Helen obtained a B.A. degree from a good college. . . ."[14] Since these are exactly the kinds of health histories and educational opportunities that comprise environmental differences, especially those between Negroes and whites, they should be regarded as supporting an environmentalist view of racial differences in IQ. The Gladys-Helen IQ difference is rare; the Gladys-Helen environmental difference is also rare. To have both rare events occurring together is evidence not of randomness but of strong nonchance association.

RELEVANCE TO RACE DIFFERENCES

How is all of the foregoing related to the problem of interpreting the 15-point difference between the mean IQs of black and white

[14] Jensen, *op. cit.*, 1970, p. 142.

Americans? The heritability of IQ among whites is one thing, but extending the conclusions to Negroes in the United States is another. Herrnstein put it this way: "Although there are scraps of evidence for a genetic component in the black-white difference, the overwhelming case is for believing that American blacks have been at an environmental disadvantage." [15]

We have seen that, given large enough environmental differences, identical twins reared apart can also exhibit IQ differences of 15 points or more. The question thus becomes: Are the environmental differences between Negroes and whites in the United States also large enough to produce average IQ differences of about 15 points?

The answer is that no one knows. We have not been given the necessary information about the way in which environment differences were measured in the studies by Newman et al. and Burt. Even if we had much more detailed information of this kind, we would then need to apply the same or comparable techniques to the measurement of the environments of representative samples of Negroes and whites in the United States.

Suppose such measurements revealed differences with magnitudes like those experienced by the identical twins represented by dots in the upper right-hand corner of figure 1 or by dots of the same kind in scatterplots based on Burt's data. Then we would have evidence that Negroes and whites would differ in IQ about as much as they do even if they had identical kinds of IQ-determining genetic compositions.

Although we lack the data necessary for environment comparisons of this kind, we can offer some reasonable conjectures based on the history of the two races, especially the blacks, in the United States. American history since 1700 has designed and executed a massive experiment in which radical manipulations of the environment constituted the experimental treatment. One substantial fraction of the population was enslaved, literally, not figuratively.

Then, after being freed, it was subjected to an elaborate, pervasive, systematic, and rigorously enforced set of social, political, economic, and educational discriminations. The treatment operated so as to impair the fabric of that fraction's familial and educational life. The experimental group was deprived of books and access to opportunities to hear standard English. Its workers were kept so physically tired by hard labor that they seldom could find energy for self-educative activities demanding intellectual effort. The experimental fraction was insulted, impoverished, made fearful, and instilled with self-hatred. In short, it would be difficult for psychologists, using what research has yielded concerning factors affecting cognitive functioning and development, to plan an environment better designed to harm the average intelligence of an experimental group consisting of about a 10 percent sample of the nation's population.

Unfortunately for the validity of the experiment, its design had a basic flaw: The subjects were not randomly assigned to the alternative treatments. Instead, the experimental treatment was confounded with the variable of race. All members of the experimental group were Negroes. Hence it has been impossible to determine, at least on the

---

[15] Richard Herrnstein, "IQ," *The Atlantic*, October, 1971, p. 57.

basis of the logic of experimental design, whether the resulting differences in the numerous dependent variables, including IQ, should be attributed to the treatment or to the race of the subjects.

Inasmuch as the experiment was flawed, it would seem to be bad research work to compound that error with another. If we cannot be sure that the educational and economic inequalities of Negroes result from the grievous experimental treatment to which they have been subjected, should we leap to the conclusion that it was their genetic makeup?

For it should be recalled that all of the research on identical twins reared apart, and almost all of that on the other relationship and rearing combinations, has been done with white subjects only. Hence that research has dealt with only the environmental variances to which whites are subjected. We have no way of being sure that those environmental variances have been large enough to embrace, at the low end of the scale, the environments to which Negroes have been subjected. It thus becomes dubious in the extreme to conclude that the IQ heritability values found for whites only would be found for whites and blacks together. And it becomes correspondingly dubious, therefore, to conclude that Negroes cannot be helped, through better education, to achieve educational and economic equality.

Yet such a conclusion seems to be implicit in suggestions that eugenic or birth-deterrent measures should be taken to reduce these inequalities. Persons with low IQs are born ineducable and unemployable, it is reasoned, and hence become burdens to themselves and society. So it is better to keep them from being born. Just as couples likely to produce physically nonviable or seriously defective children are cautioned against having them, so persons with low IQs should be discouraged, perhaps with monetary compensations, from having children.

How to reply to such a proposal? We simply do not know enough, and are much too unsure about what knowledge we do have, to establish an adequate scientific and moral basis for such a social policy. But we do know that IQ is not everything, in our society or any other. Other human qualities, as Michael Young's *Rise of the Meritocracy* made clear, are equally valuable to us and our fellow men. There are many special intellectual abilities, artistic and musical talents, and types of creativity that are missed by IQ tests.

Also, the IQs of children are not predictable enough from those of their parents to justify preventing anyone other than the most defective from having children. And such proposals overlook the regression effect which occurs when correlations are less than perfect. Such an effect makes extreme values of one variable go along with less extreme values of another variable imperfectly correlated with it. Regression effects, relating both the empirical and the statistical facts inherent in correlations of .5 between the IQs of parent and children,[16] make parents with IQs below 100 likely to have children with IQs closer to 100.

Finally, it seems likely that such a policy would drain resources away from educational and other efforts aimed at improving environmental influences on educational achievement and employability.

---

[16] L. Erlenmeyer-Kimling and L. F. Jarvik, "Genetics and Intelligence: A Review," *Science*, 1963, pp. 1477–79.

BETTER EDUCATIONAL R. & D. NEEDED

Educational and other environment-improvement approaches to the problems of racial inequality have not been given anything close to an adequate trial thus far in the United States. The doubts often expressed concerning the outcomes of Project Headstart are based on disregard of the definite evidence of positive results in substantial subsamples.[17] The pessimists also forget that the Headstart projects evaluated were begun in a hurry and so were inadequately planned. They were handicapped by makeshift staff and curriculum materials.

What would come from adequately planned and stably supported projects? If based on systematically researched experimental alternatives, such projects would produce, in the opinion of many behavioral scientists, solid and worthwhile improvements in the educational achievement of children from low-income homes. It is *not* true that "compensatory education has been tried and it apparently has failed." Compensatory education needs more research and better supported tryouts over a period of decades, not merely a single enthusiastic presidential administration. Its support needs to be made independent of partisan politics, at least as much as cancer and weapons research have been freed in this way. It needs better-trained and larger research forces, competent in the best methods and theories of the behavioral and, where relevant, the biological and physical sciences. It needs research not just on what and how to teach but on how to train teachers and the administrators of their schools. In short, the effort to eliminate racial inequality in American society needs better education and the research to make that better education possible.

---

[17] Westinghouse Learning Corp. and Ohio University, "The Impact of Headstart: An Evaluation of the Effects of Headstart on Children's Cognitive and Affective Development." Presented to the Office of Economic Opportunity, Washington, D.C., 1969.

AN EXPERIMENT IN THE PREVENTION OF CULTURAL-FAMILIAL MENTAL RETARDATION [1, 2, 3]

BY RICK HEBER, PH. D., DIRECTOR

AND

HOWARD GARBER, PH. D., RESEARCH ASSOCIATE

REHABILITATION RESEARCH AND TRAINING CENTER IN MENTAL RETARDATION

University of Wisconsin

Madison

LIST OF TABLES AND FIGURES

In the United States it is generally estimated that 3 percent of the population may be mentally retarded. This amounts to more than 6 million persons. However, in about 80 percent of this number, there is no identifiable gross pathology of the central nervous system. Furthermore, this group of the mentally retarded, without identifiable pathology of the central nervous system, is almost exclusively found

[1] Research supported in part by grant 16-P-56811/5-06, formerly RT-11, from the Rehabilitation Services Administration of the Social and Rehabilitation Service of the Department of Health, Education, and Welfare.

[2] Presented at Warsaw, Poland.

[3] Proceedings of the Second Congress of International Association for the Scientific Study of Mental Deficiency, August 25–September 2, 1970.

among the populations of economically distressed urban and rural areas.

Dr. Seidenfeld will discuss the recent advent of social concern and the awakening of conscience concerning America's "poor," and the concern for her racial and ethnic minorities who are so heavily represented among the "poverty stricken." With this stirring of conscience, there has been increasing acceptance, and little critical challenge, of the view that the high frequency of mental retardation found among the "poverty stricken" is directly attributable to deprivation of opportunities available to the "nonpoor" to learn and practice intellectual skills.

However, it should be obvious to all that simple awareness of the high frequency of mental retardation in areas where the economically or otherwise disadvantaged are concentrated, is not sufficient to conclude that social deprivation in the "slum" environment, in any general sense, causes the retardation enountered there. Such a generalization ignores the fact that most children reared by economically disadvantaged families are by no means mentally retarded. In fact, a majority of children reared in the slums of the cities grow and develop and learn relatively normally in the intellectual sense.

In order to learn more about the distribution of mental retardation in the population of a city slum, we have, over the past few years, conducted a series of surveys. Our survey area is a residential section of Milwaukee (a city of 800,000), which is characterized by census data as having the lowest median family income, the greatest population density per living unit, and the greatest rate of dilapidated housing in the city. For the United States, it is a typical urban slum and yields by far the highest prevalence of mental retardation among school children in the city. In our first survey, all families residing in this slum who had a newborn infant and at least one other child of the age of 6, were selected for study.

TABLE I.—*Distribution of child IQs as a function of maternal intelligence*

Mother's IQ	Percent of mothers	Children's IQ: Greater than 90 percent	Children's IQ: 80 to 90 percent	Children's IQ: Less than 80 percent
Greater than 80	54. 6	65. 8	47. 3	21. 9
Less than 80	45. 4	34. 1	52. 7	78. 2

The major survey finding of relevance to this discussion is that the variable maternal intelligence proved to be the best single predictor of the level and character of intellectual development in the offspring. Mothers with IQs less than 80, although comprising less than half the total group of mothers, accounted for almost four-fifths of the children with IQs below 80 (see table 1).

It is generally acknowledged that the "slum-dwelling" children score lower on intelligence tests as they grow older. However, as can be seen here, in figure 1, the mean measured intelligence of offspring of mothers with IQs above 80 is relatively constant. And it is only the

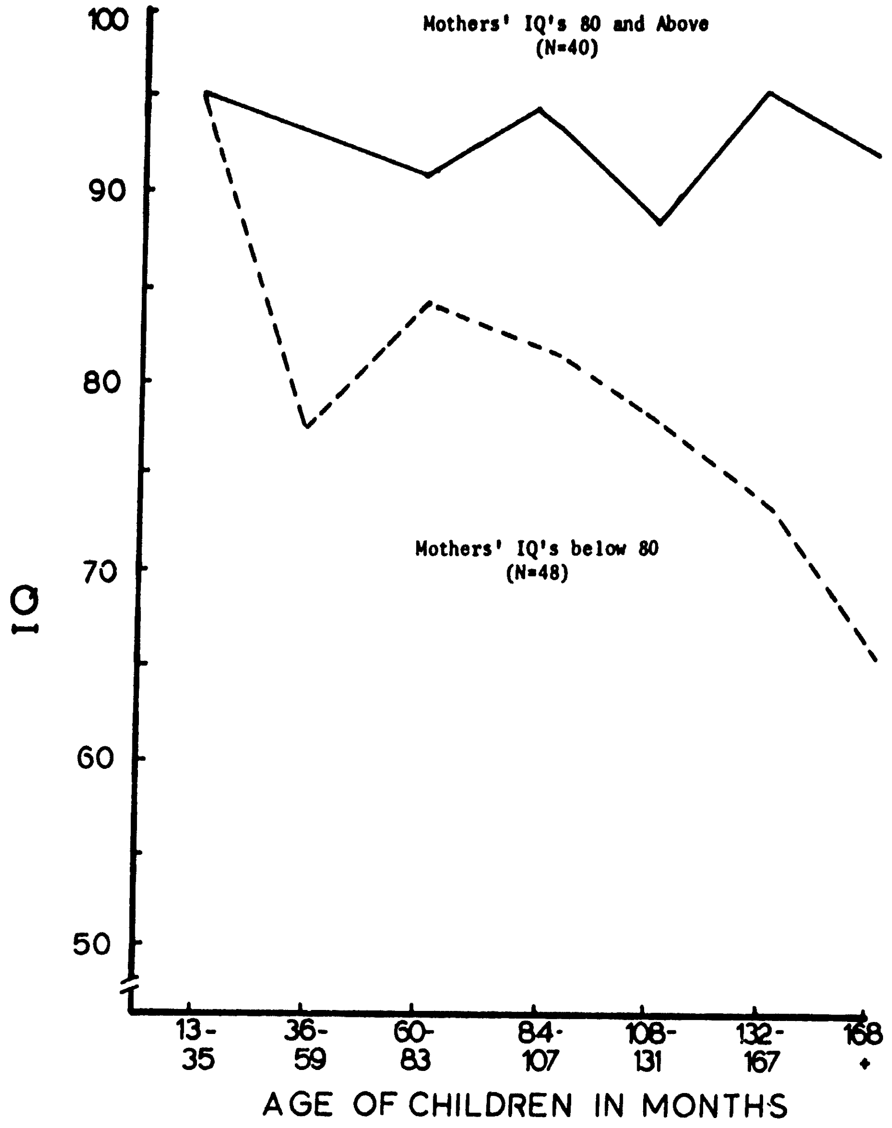


Figure 1. IQ Decrements in Disadvantaged Children Whose Mothers are Mentally Retarded

Source: Heber, R. F., Dever, R. B. and Conry, J. The influence of environmental and genetic variables on intellectual development. In H. J. Prehm, L. A. Hamerlynck and J. E. Crosson (Eds.), Behavioral Research in Mental Retardation (Eugene, Oregon: University of Oregon, 1968), p. 9.

children of mothers with IQs below 80 who show a progressive decline in mean intelligence as age increases.

Further, the survey data showed that the lower the maternal IQ, the greater the probability of offspring scoring low on intelligence tests. For example, as you can see in table 2, the mother with an IQ below 67 had a roughly, fourteenfold increase in the probability of having a child test below IQ 67 as compared with the mother whose IQ fell at or above 100.

TABLE II.—*Probability of child IQ following within IQ ranges as a function of maternal IQ*

Child's IQ	Mother's IQ			
	Greater than 100	84 to 99	68 to 83	52 to 67
Greater than 100	1	0. 98	0. 67	0. 25
84 to 99	1	1. 02	. 95	. 93
68 to 83	1	1. 57	1. 24	2. 20
52 to 67	1	2. 36	3. 70	14. 20

In our original survey, the fathers were not evaluated. However, in a second survey of 519 consecutive newborns in our study area, fathers, mothers, and all children over the age of 2 were administered the PPVT of intelligence. The results of this survey added further to our knowledge of the distribution of intellectual functioning within a "slum" population. First, there was a rather striking congruence of maternal and paternal IQ. Of mothers below IQ 70, 65 percent had husbands who also scored below 70, and only 14 percent had husbands who scored above 100. By contrast, not a single mother scoring above IQ 100 had a husband who scored below IQ 80. Further, there were approximately twice as many mothers under 20 and over 35 in the below 70 IQ group as compared with mothers above 100. This was reflected in a substantially greater number of offspring in families where both mother and father tested below 70. There was an average difference of 1.2 children between these families and those where the mother and father both tested above 100. Considering that these families were estimated to be, on the average, about halfway through their childbearing years, the mean difference in the number of offspring in completed families might be on the order of 2½. The adverse consequences of this differential in reproductive activity are, of course, of great social concern, irrespective of one's views of the etiology of the intellectual deficiency in the parents.

These surveys have convinced us that the very high prevalence of mental retardation associated with the "slums" of American cities is not randomly distributed but, rather, is strikingly concentrated within individual families who can be identified on the basis of maternal intelligence. In other words, the source of the excess prevalence of mental retardation appears to be the retarded parent residing in the "slum" environment, rather than the "slum" itself in any general sense.

At first glance, these population survey data seem to suggest support for the prepotence of hereditary determinants of "cultural-familial" mental retardation. However, simple casual observation suggested that the mentally retarded mother residing in the "slum" creates a social environment for her offspring which is distinctly different from that created by the "slumdwelling" mother of normal intelligence. As a result, we have been pursuing a longitudinal, prospective investigation designed to contribute to our understanding of the determinants of the kind of retardation which perpetuates itself from parent to child in the "slum-dwelling" family.

We are proceeding to test the social deprivation hypothesis by a methodology which we hope will enable us to determine whether the development of intellectual deficiency may be prevented (as opposed to cured or remediated) by displacing the presumed adverse or negative factors in the social environment of disadvantaged children who become retarded. The results of our population survey data provided us with the ability to initiate a longitudinal research program by providing us with the ability to select a sample which would be small enough for practical experimental manipulation, but yet which would yield a sufficient number of cases who would later become identifiable as mentally retarded.

As a consequence of the survey data, we have utilized maternal IQ as a basis for selection of a group of newborns, with confidence that a substantial percentage would be identified as mentally retarded as they grow older. By screening all mothers of babies born in our survey area over a period of about 1 year, we identified mothers of newborns with IQs less than 70. We have drawn 40 of these mentally retarded mothers from the subject pool and assigned them randomly to either our experimental or our control group. Beginning in the first few weeks of life, we have undertaken a comprehensive intervention in the social environment of our experimental newborns. The objective of this intervention was to displace all of the presumed negative factors in the social environment of the infant being reared in the slum by a mother who is herself retarded. We are, thereby, testing the "social deprivation" hypothesis of etiology by seeing whether it is possible to prevent retardation from occurring in the offspring of these retarded mothers.

Should the experimental children reach the age of 6 or 7 and exhibit normal intelligence, we will know that it is possible, through our experimental program, to prevent mental retardation from occurring at the present high frequency in this group. Should they exhibit a retarded level of functioning, we will know that their intensive exposure to learning experiences was not sufficient to displace their genetic predispositions for intellectual functioning.

The intervention was initiated in the home shortly after the mother returned from the hospital. As soon as a feeling of trust on the part of the mother was achieved, her infant was introduced to our Infant Education Center and she was exposed to a maternal rehabilitation program. The program for the retarded mothers was designed to modify those aspects of the environment which the mother herself creates or controls. These maternal rehabilitation services are in the

form of occupational training for the mother as well as training in homemaking and baby-care techniques.

At the Infant Education Center, the infants receive a customized, precisely structured program of stimulation. The infants are picked up in their homes early each morning by their infant teachers and are transported to the center where they remain until late afternoon. Infant stimulation teachers follow an intensive program which has been prescribed in detail. Essentially, it includes every aspect of sensory and language stimulation which we believe may have relevance for the development of intellectual abilities. Its major emphasis are efforts designed to facilitate achievement motivation, problem-solving skills, and language development.

In order to assess the effects of this kind of comprehensive intervention into the natural environment of the infant reared by a retarded mother, we are undertaking an intensive schedule of measurements. These include, in addition to the standardized tests of development and intelligence, a number of experimental measures of learning and performance, and a variety of measures of language development. Because of the dictates of our sample selection procedure, there is an age range in our infants of about 1 year, with the oldest of our experimental and control children now being about 4 years of age.

Because of our time limitation this morning, we can do no more than present findings on a few measures which illustrate the trends in our data. Figure 2 shows the performance of experimental and control groups on an experimental task designed to assess the response strategy employed by the child. It is a matching task in which the child may respond correctly according to color or form. He cannot respond correctly if he follows any other strategy such as position responding or alternation responding. In this case, responding to the dimension of either color or form is more developmentally advanced than the failure to adopt a dimensional strategy. As you can see here, none of the younger controls and only 30 percent of the older controls showed a dimensional response; that is, in terms of color or form. By contrast, 60 percent of the younger experimentals and 100 percent of the older experimentals showed undimensional responding.

Figure 3 shows performance on another experimental task, which also assesses dimensional responding, but it is more difficult, in that it is a sorting, rather than matching, task. Here, the younger groups were equivalent in terms of dimensional responding; however, the older experimental group showed 91 percent dimensional responding as opposed to 50 percent for the control group.

Figure 4 shows the tendency of experimental and control children to perseverate by continuing to respond to the stimulus, on a discrimination learning task, which has been shifted from the appropriate to the inappropriate choice. For the experimentals, the percentage of children showing a perseverative response decreased from 50 percent for the younger group to 19 percent for the older group. By contrast, the controls actually increase their percentage of perseverative responding from 66 percent at the younger level to 100 percent at the older level.

DIMENSIONAL PREFERENCE IN MATCHING

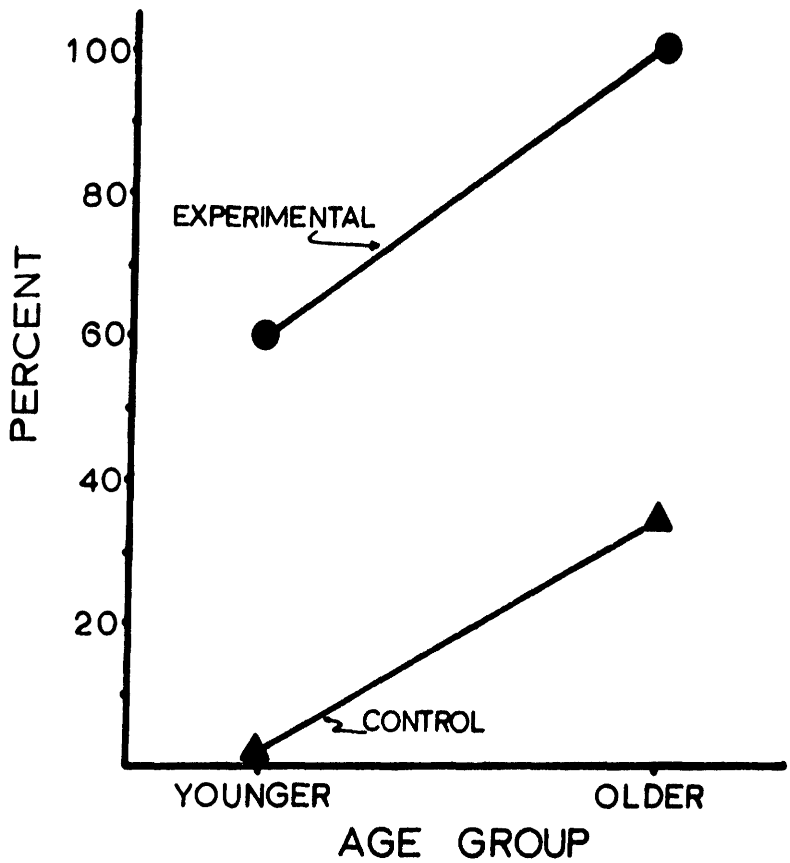


Figure 2.

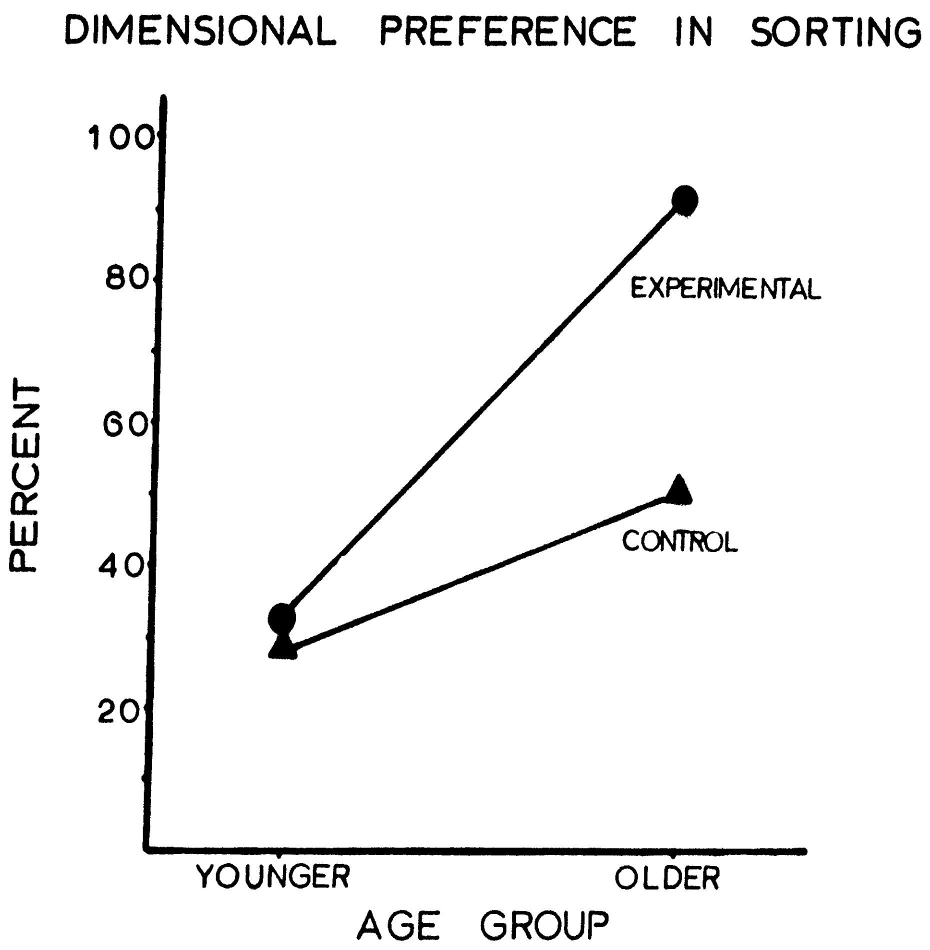

DIMENSIONAL PREFERENCE IN SORTING
100
80
60
40
20
PERCENT
EXPERIMENTAL
CONTROL
YOUNGER
OLDER
AGE GROUP

Figure 3.

PERSEVERATION IN PROBABILITY TASK

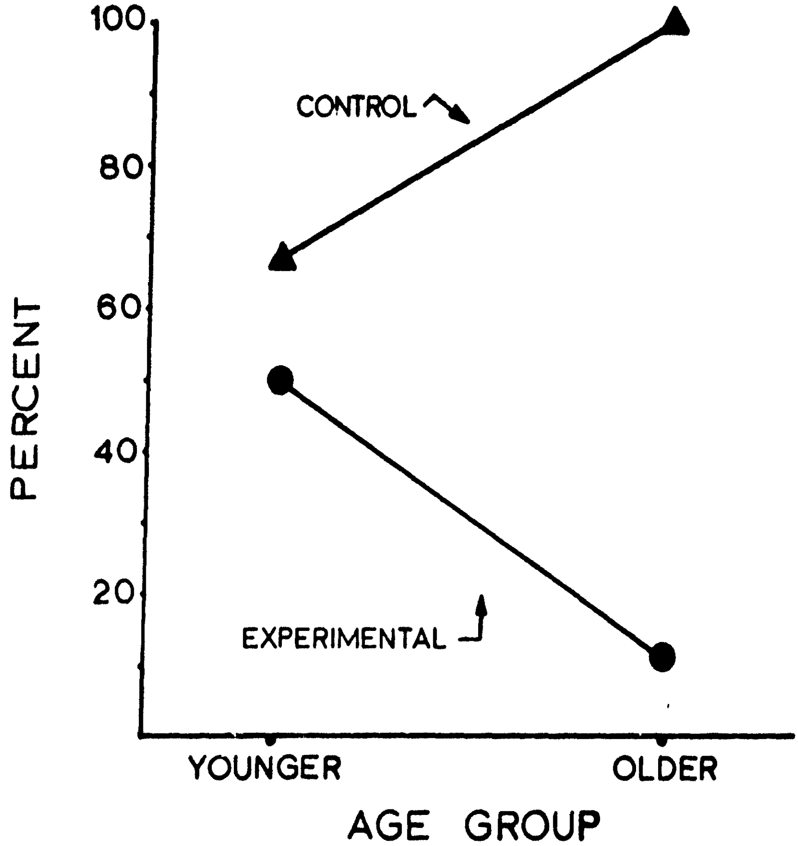


Figure 4.

The most striking differences in the performance of the experimental and control children are reflected in the measures of language performance. Beginning at 18 months, free speech samples and formal language tests have been administered every 6 weeks.

TABLE III.—*Quantitative analysis of language sample: scores indicate number of total vocabulary per 10 minute unit of time for the experimental and control groups*

Age at sample	19 months	22 months	25 months	28 months
Experimental group:				
$\bar{X}$ (including no vocabulary)..	4.00	14.86	23.69	20.91
$\bar{X}$ (excluding no vocabulary)..	6.66	14.86	26.32	20.91
Control group:				
$\bar{X}$ (including no vocabulary)..	3.30	3.81	2.31	9.91
$\bar{X}$ (including no vacabulary)..	8.25	5.44	3.70	17.35

Figure 5 illustrates the total number of different words spoken (per 10 minutes time unit) for free speech samples between 19 and 28 months (see table 3). The solid lines include the zero scores for those infants who did not vocalize during the free speech sample. The dotted lines exclude the zero scores. Note the marked spurt in vocabulary production for the experimental group between 19 and 25 months, a spurt which does not begin to occur until 28 months for the control group. And note also the number of zero scores which continue to occur at 28 months for the control group. The slight decrease in vocabulary production at 28 months for the experimental group can be accounted for by their beginning to concentrate on the acquisition of grammatical structure.

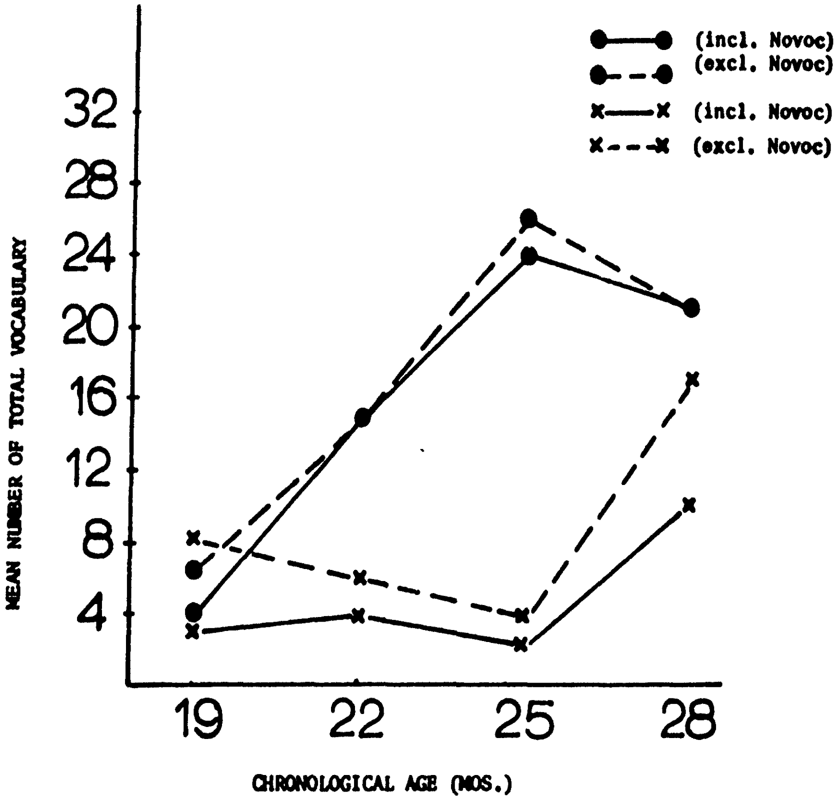
Figure 6 shows the results of a sentence repetition test administered between 36 and 45 months. This test requires the child to repeat 34 sentences which vary in length and in grammatical complexity. As you can see, there is a substantially greater number of words omitted at all age levels for the control children. For words substituted incorrectly, the control group surpasses the experimental group at 36 months, but at 42 months, there are distinctly fewer substitutions made by the experimental group.

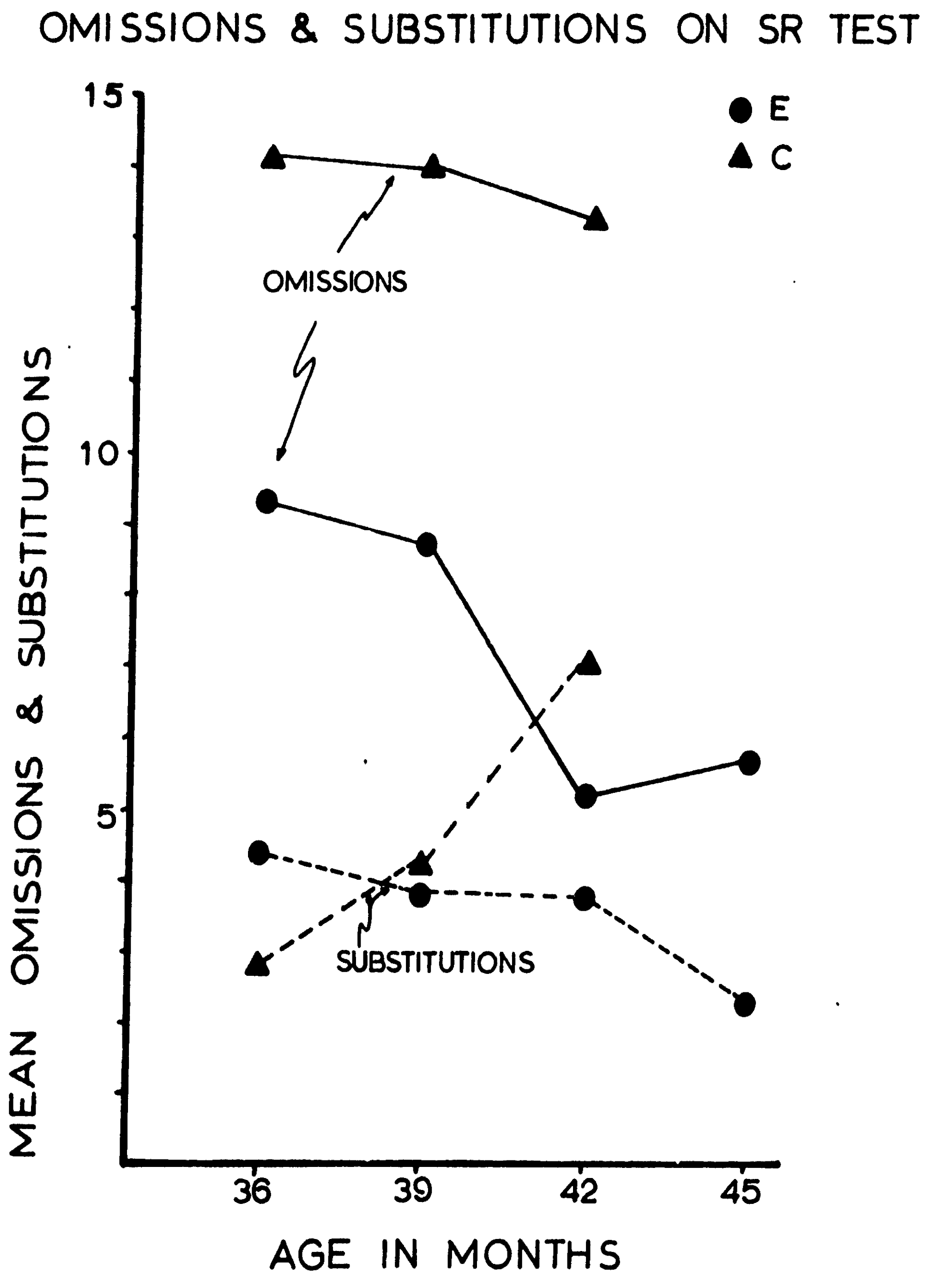
Figure 7 shows the number of exact repetitions by the experimental and control groups. The experimental group is distinctly superior to the control group at all age levels.

A test of grammatical comprehension has also been given at 3 month intervals beginning at 36 months. This test assesses the child's comprehension of 16 different grammatical features or rules of the English language. Figure 8 shows the percentage of sections on the test performed correctly by each group. Each section of the test refers to one feature of language. Again, the difference in favor of the experimental group is marked.

Figure 9 presents Cattel and Binet IQ data, from 24 months on, for the experimental and control groups and for a contrast group. In effect, it summarizes the present differential in development between the experimental and control groups. The dotted line represents the

Figure 5. Quantitative Analysis of Language Sample: Scores Indicate Number of Total Vocabulary Per Ten Minute Unit of Time for the Experimental and Control Groups





OMISSIONS & SUBSTITUTIONS ON SR TEST
15
10
5
MEAN OMISSIONS & SUBSTITUTIONS
E
C
OMISSIONS
SUBSTITUTIONS
36
39
42
45
AGE IN MONTHS

Figure 6.

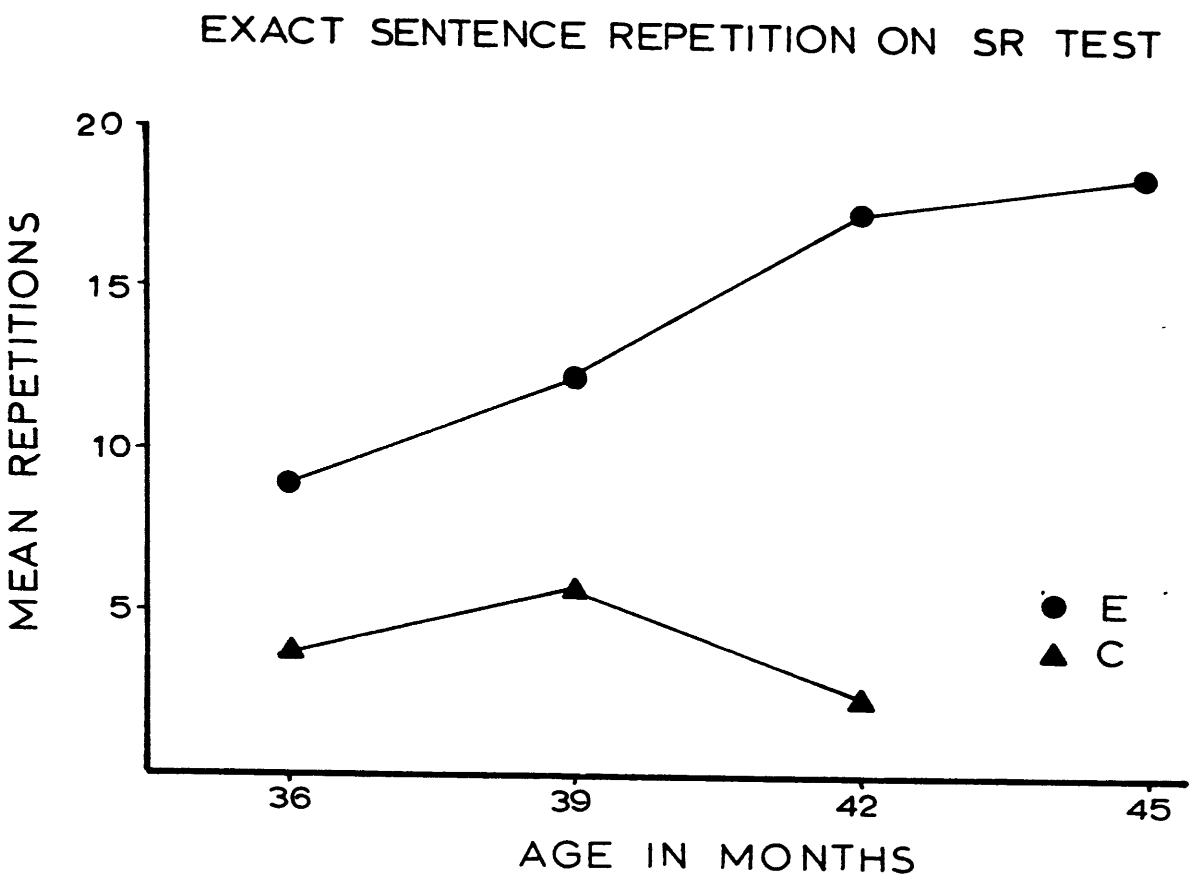

EXACT SENTENCE REPETITION ON SR TEST
MEAN REPETITIONS
20
15
10
5
36
39
42
45
AGE IN MONTHS
E
C

Figure 7.

PERCENTAGE OF SECTIONS CORRECT ON <u>G-C</u> TEST

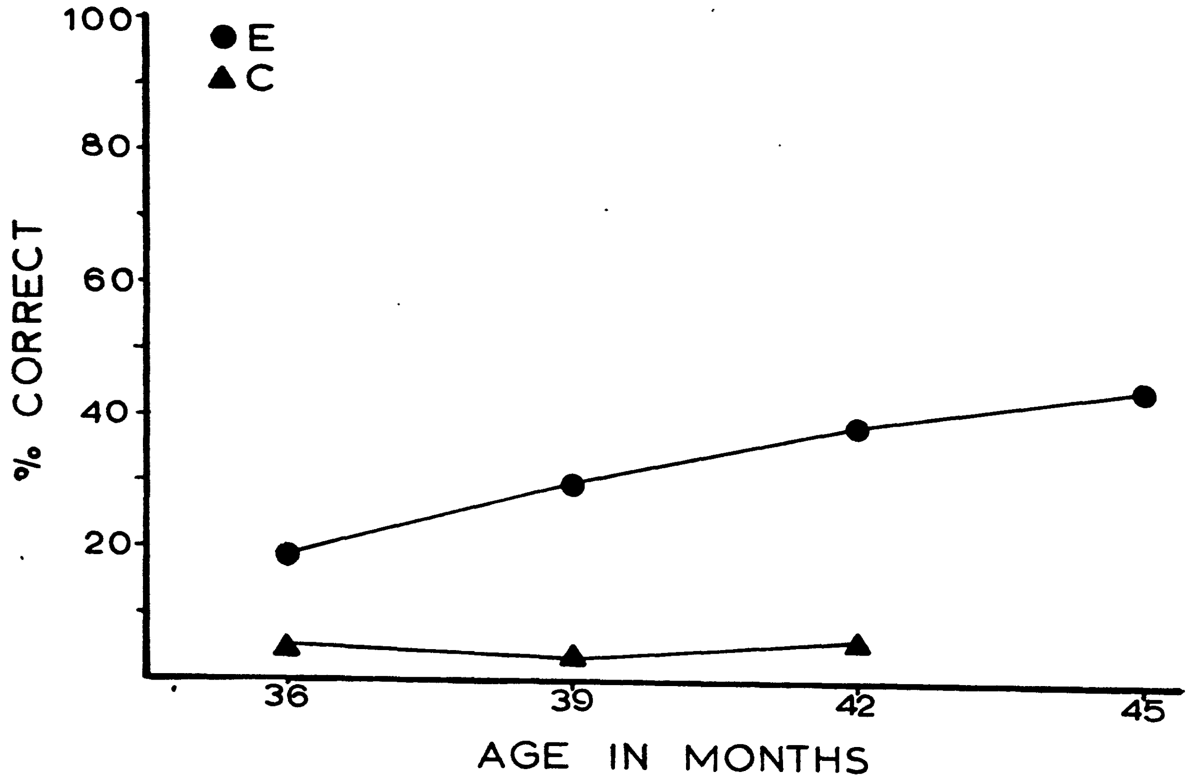


Figure 8.

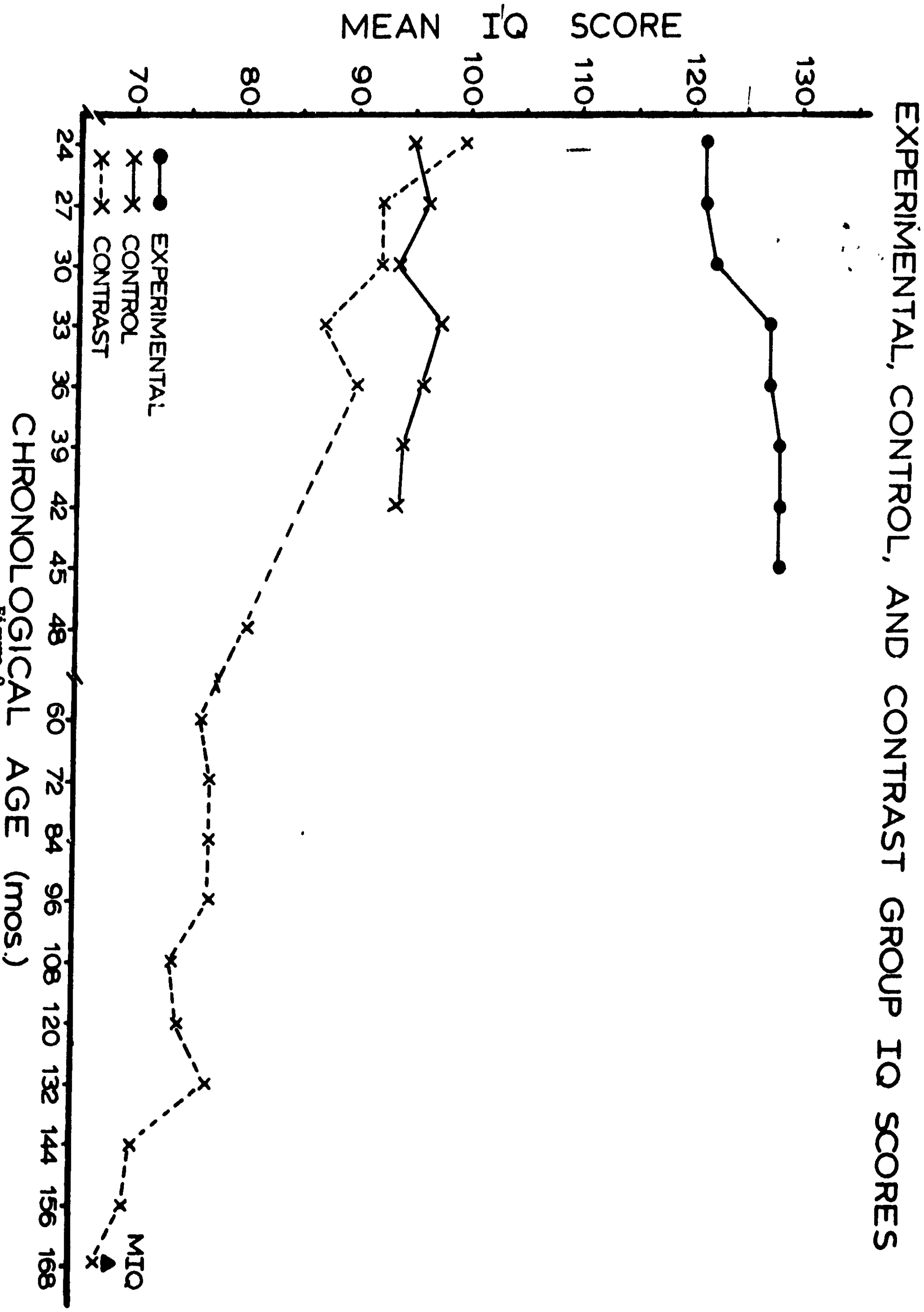

EXPERIMENTAL, CONTROL, AND CONTRAST GROUP IQ SCORES
MEAN IQ SCORE
70
80
90
100
110
120
130
24 27 30 33 36 39 42 45 48 60 72 84 96 108 120 132 144 156 168
CHRONOLOGICAL AGE (mos.)
EXPERIMENTAL
CONTROL
CONTRAST
MIQ

Figure 9.

mean IQs of offspring of mothers with IQs below 75, taken from our original population survey. It depicts the patterns of development expected for our actual control group. You can note that the pattern of performance of our control group is not discrepant from the contrast group.

You will recall that our hypothesis was in terms of preventing the relative decline in development of the experimental group which we see in the contrast group and which we expect to see in the control group. We did not anticipate the marked acceleration which you see in the experimental group. At 42 months, the discrepancy between the experimental and control group is on the order of 33 IQ points.

Our awareness of the numerous pitfalls and hazards of infant measurement leads us to extreme caution in interpretation of our present data. Our experimental infants have obviously been trained in skills sampled by the tests and the repeated measurements have made them test-wise. They have been provided with intensive training to which no comparable group of infants has ever been exposed, to the best of our knowledge. Have we, thereby, simply given them an opportunity to learn and practice certain intellectual skills at an earlier age than is generally true? And if so, will their apparent acceleration in development diminish as they grow older?

Nevertheless, the performance of our experimental children, today, is such that it is difficult to conceive of their ever being comparable to the "lagging" control group. We have seen a capacity for learning on the part of extremely young children surpassing anything which, previously, I would have believed possible. The trend of our present data does engender the hope that it may prove to be possible to prevent the kind of mental retardation which occurs in children reared by parents who are both poor and of limited ability.

77-942 O—72——32

[From the Atlantic Monthly, September 1971]

I.Q.

By Richard Herrnstein

1. If differences in mental abilities are inherited; and
2. If success requires those abilities; and
3. If earnings and prestige depend on success;
4. Then social standing will be based to some extent on inherited differences among people.

False ☐ True ☐

The measurement of intelligence forced its way into America's public consciousness during World War I, when almost 2 million soldiers were tested by the Army and categorized as "alpha" and "beta," for literates and illiterates respectively. The lasting effect of that innovation has not been the surprise at learning that the average American soldier had an intelligence equal to that of a 30-year-old, or that artillery officers were substantially brighter than medical officers, or any of the myriad other statistical curiosities. Even if those facts are still as true as they were in 1918, the lasting effect has been the mere use of the tests and their serious consideration by responsible people. For intelligence tests, and the related aptitude tests, have more and more become society's instrument for the selection of human resources. Not only for the military, but for schools from secondary to professional, for industry, and for civil service, objective tests have cut away the traditional grounds for selection—family, social class, and, most important, money. The traditional grounds are, of course, not entirely gone, and some social critics wonder if they do not lurk surreptitiously behind the scenes in our definition of mental ability. But at least on the face of it there is a powerful trend toward "meritocracy"—the advancement of people on the basis of ability, either potential or fulfilled, measured objectively.

Lately though, the trend has been deplored, often by the very people most likely to reap the benefits of measured intellectual superiority. More than a few college professors and admissions boards and even professional testers have publicly condemned mental testing as the basis for selection of people for schools or jobs. The IQ test, it is said with fervor, is used by the establishment to promote its own goals and to hold down the downtrodden—those nonestablishment races and cultures whose interests and talents are not fairly credited by intelligence tests. These dissenting professors and testers are naturally joined by spokesmen for the disadvantaged groups. We should, these voices say, broaden the range of humanity in our colleges (to pick up the most frequent target) by admitting students whose low college entrance examination scores might otherwise have barred the way. For if the examinations merely fortify an arbitrarily privileged elite

in its conflict with outsiders, we must relinquish them. The ideals of equality and fraternity must, according to this view, take precedence over the self-interest of the American-Western European middle class.

The issue is intensely emotional. It is almost impossible for people to disagree about the pros and cons of intelligence testing and long avoid the swapping of oaths and anathema. Yet should not the pros and cons be drawn from facts and reason rather than labels and insults? For example, is it true that intelligence tests embody only the crass interests of middle America, or do they draw on deeper human qualities? Is the IQ a measure of inborn ability, or is it the outcome of experience and learning? Can we tell if there are ethnic and racial differences in intelligence, and if so, whether they depend upon nature or nurture? Is there only one kind of intelligence, or are there many, and if more than one, what are the relations among them? If the tests are inadequate—let us say, because they overlook certain abilities or because they embody arbitrary cultural values—how can they be improved? For those who have lately gotten their information about testing from the popular press, it may come as a surprise that these hard questions are neither unanswerable nor, in some cases, unanswered. The measurement of intelligence is psychology's most telling accomplishment to date. Without intending to belittle other psychological ventures, it may be fairly said that nowhere else—not in psychotherapy, educational reform, or consumer research—has there arisen so potent an instrument as the objective measure of intelligence. No doubt intelligence testing is imperfect, and may even be in some sense imperfectible, but there has already been too much success for it to be repudiated on technical grounds alone. If intelligence testing is to change, it must change in light of what is known, and more is known than most might think.

Mental testing was one of many responses within psychology to Darwin's theory of evolution. In fact, the connection here is intimate and direct, for the idea of measuring mental ability objectively was first set forth by Francis Galton, the younger cousin of Charles Darwin. Far more versatile (perhaps smarter) than his great cousin, Galton was a geographer, explorer, journalist, mathematician, eugenicist (he coined the term), and articulate essayist. In 1869, just a decade after Darwin launched modern biology with the "Origin of Species," Galton published "Hereditary Genius," which applied evolutionary thinking to the question of intellect. Galton noted, first, that men varied greatly in their intellectual capacity and, second, that various kinds of excellence run in families, suggesting that the basis of intelligence may be inherited. Going back through British history, Galton found that judges, statesmen, prime ministers, scientists, poets, even outstanding wrestlers and oarsmen tended, for each kind of endeavor, to be related by blood. The eminent families of Great Britain were taken as evidence of superior human strains, comparable to the natural biological variations that figure so prominently in the doctrine of evolution. Today, our sensitivity to the role of the environment (not to mention such mundane complications as money and family connections) make us skeptical of his evidence. Nevertheless in the first flush of Darwinian social theorizing, Galton called for constructive change. The inheritance of human capacity implied "the practicability of supplanting inefficient human stock by better strains," and led him "to consider whether it might not be our duty to do so by such efforts as may

be reasonable, thus exerting ourselves to further the ends of evolution more rapidly and with less distress than if events were left to their own course."

Galton was not much more content with the genealogical approach to mental ability than are we today. Within a few years, he was trying to test mental ability directly, but the problem was how to do it. In 1882, Galton set up a small laboratory, in a London museum where people could, for a fee, have their hearing, vision, and other senses tested. Galton knew that mental defectives—idiots and imbeciles—often lacked sensory acuity, and he guessed that there might be a reasonably consistent relation between intelligence and sensory keenness in general. As it turned out, his hunch was wrong, or at least not right enough to be useful as a way of testing on a large scale.

Galton was soon just one of many scientists searching for a practical intelligence test, with no one much worried at this point about the ultimate definition of intelligence. Intuition and commonsense set the standards as the few simple measures of sensory acuity gave way to a host of tests, some sensory and others drawing on other psychological processes. An American pychologist named James McK. Cattell coined the phrase "mental test" in 1890 in an article recounting his studies at the University of Pennsylvania on the mental abilities of students. In addition to simple sensory function, Cattell measured color discrimination, time perception, accuracy of hand movement, and memory; and he collected descriptions of imagery. People no doubt differed, but it was hard to know what to make of the differences. By the mid-1890s, testing had attracted so much attention that professional organizations began taking note of it. The newly founded American Psychological Association formed a committee in 1895 "to consider the feasibility of cooperation among the various psychological laboratories in the collection of mental and physical statistics"; in 1896 the American Association for the Advancement of Science instructed a committee of its own "to organize an ethnographic survey of the white race in the United States." The quotations in both cases are Professor Cattell's words; he was a member of both committees and was determined that the ethnographic survey for AAAS include some of APA's mental (and physical) tests.

For all of the ferment, it was not yet certain that anything useful was brewing. There was spirit and energy in abundance, but there were as yet no indisputably good tests. It took the work of a French psychologist named Alfred Binet to make intelligence testing practical. In a key article written in 1895, Binet and his junior collaborator, Victor Henri, argued for mental testing based not on sensory or motor functions but on the psychological processes thought to be involved in intelligence. Instead of supposing that being smart is the outcome of having keen senses or speedy reactions, Binet argued that intelligence operates at its own level and that, therefore, a proper test must engage the person at that very level. As for what such tests might be, Binet, like everyone else in 1895, was just guessing. The article suggested a variety: tests of memory, mental imagery, imagination, attentiveness, mechanical and verbal comprehension, suggestibility, aesthetic appreciation, moral sensibility, the capacity to sustain muscular effort, and visual judgment of distance.

Binet criticized his contemporaries for their preoccupation with sensory and other simple processes, which, although fulfilling their desire for exactitude in measurement, had sacrificed the still more salient need for relevance. For Binet, exactitude was secondary. His pragmatism directed him to tests that sorted people out—for, whatever intelligence is, it varies from person to person. The sensory data did not distinguish among people as sharply as intuition required for a test of intelligence. Binet committed himself to seeking the tests that would do so, which was an undertaking that occupied the rest of his life. In the following 10 years, Binet and his collaborators worked on mental testing at the psychological laboratory of the Sorbonne, using as their subjects mainly children from the schools of Paris and its suburbs.

The use of children was a happy accident, for it focused attention on the chronology of intelligence. Of all the countless ways one may want to distinguish between smarter and duller people, it may not seem especially insightful to choose the simple fact that during the first 15 or so years of life, age confers intelligence (on the average). Thus, if an intellectual task sorted children according to their age, then it might properly be included in an intelligence test. In one experiment, for example, Binet tested over 500 school children by reading them a sentence and then asking them to write down as much of the sentence as they could remember. Between the ages of 9 and 12 (the ages tested), each successive grade of student did better, albeit slightly, than the grade younger. From this, Binet knew that the "sentence-reproduction test" could be taken as one measure of mental capacity. And knowing that, he could say that if two children of equal age differed in their sentence-reproduction scores, they were to some degree different in intelligence. One such test was, however, far from a usable measure of general intelligence, as Binet well knew.

As the years passed, Binet and others stocked a rich store of norms and measures of mental ability, based on many tests of many children. Even Binet's own two daughters were the subject of intensive study, culminating in a book called "The Experimental Study of Intelligence" (1902), in which the vital psychological facts about the teenage girls were expressed as scores on their father's tests of word-writing speed, mental imagery, sentence completion, and so on. It was to Binet, therefore, that the Minister of Public Instruction turned in the fall of 1904 when he wanted a better way to spot subnormal children in the Parisian schools. The children were to be put into special schools where they could be helped, but the first problem was to find them. If mental tests were any use at all, here was a task to prove it. Binet and his psychiatrist collaborator, Theodore Simon, decided to use a series of tests graded in difficulty, first standardized on normal children of various ages.

The idea of using equivalent age as the measure of intelligence was obvious only after Binet, not before him, for it was one of those rare and elegant turns that make for historic innovation. Here were some tests that distinguished between children of different ages, on the average. However, at each age some children did better than their exact chronological peers. Those children, he had found, were judged by teachers to be bright or gifted. Conversely, other children did

worse than their peers and were judged to be dull. Hence, if all one knows about a child is that he outperforms his age peers, he can still be assumed to be bright. If his performance matches his age, he is probably an average child in intelligence. And if he underperforms, he is probably dull. As Binet well knew, the chronological approach to intelligence finessed the weighty problem of defining intelligence itself. He had measured it without having said what it was. It took a while to know whether the sleight of hand had in fact yielded a real intelligence test or just an illusion of one.

For their first practical venture, Binet and Simon drew up a progression of 30 tests covering the range of mental capacity. At the very bottom, the examiner simply noted eye-head coordination as a lighted match was moved across the field of vision; thence he observed the making of grasping movements, imitating gestures, the following of instructions to touch various parts of the body, the naming of familiar objects, repeating sentences, arranging identical-looking objects in order of weight, constructing sentences to include three given words ("Paris," "gutter," "fortune"); and finally, the ability to distinguish between abstract words such as "liking" and "respecting." After some preliminary trials, Binet and Simon gave their test to about 50 normal children between the ages of 3 and 11, thereby establishing the cut-offs for each age. Finally, using children already diagnosed by standard clinical procedures to be idiots, imbeciles, and morons, they found the corresponding criteria for mental disability in their series of tests.

Is a retarded child really the equal of a normal child at a younger age? For example, the average 5-year-old passed the first 14 tests, while the upper limit for an imbecile was to pass the first 15 tests whatever age he was. Anyone who passed more was not an imbecile. Was Binet saying that a 12-year-old imbecile precisely equals a slightly brighter than average 5-year-old? The answer is no, for Binet specifically denied the charge. The imbecile, he said, is "infirm," the 5-year-old is healthy, and their mental processes are in some respects different, even if the difference is not captured by his test. Nevertheless, the test did its job, for a 12-year-old who tested at the 5-year-old level was, indeed, retarded, while a 5-year-old who did so was not (or at least did not seem to be at that time). As always, Binet's approach was doggedly pragmatic and empirical. He was picking out the retardates with his test more quickly, cheaply, and for all anyone knew, more accurately than ever before. The social benefits were self-evident.

The Binet-Simon test was put into use immediately and was criticized as quickly for this or that item. But criticism was corrective, for in showing that some item was not, for example, distinguishing between 3- and 4-year-olds, the critic was opening the test to improvement. An ineffective item could be dropped, a useful one added, without in the least altering the kernel idea, which was to measure intelligence by a graded series of tasks ("stunts," Binet often called them). The tests and the criticisms were rooted in actual experience with ever-growing numbers of children, adding greater and greater empirical stability to the results. In America, Great Britain, Belgium, Italy, Germany, and elsewhere, the tests were being used and perfected. In a cheering counterexample to Gresham's gloomy law, good

test items tended to drive out bad ones, and the better the test in sorting out children, the more it was used and improved. In 1908, Binet and Simon published a much-revised series of tests, to be used for rating children in general, not just retarded children. In 1911, the final Binet-Simon scale came out; it was Binet's last work, for he died that year at the age of 54. But the evolution of testing continued unabated and still does.

In the 1911 version, there were five problems which the average child of each age could or could not solve. Here, for example, are the five items for the 6-year level:

1. Distinguish between morning and afternoon.
2. Define familiar objects in terms of use.
3. Copy a diamond shape.
4. Count 13 pennies.
5. Distinguish between ugly and pretty faces.

And here are the five problems for the average 10-year-old:

1. Arrange five blocks in order of weight.
2. Draw two designs from memory.
3. Criticize absurd statements.
4. Answer comprehension questions.
5. Use three words in not more than two sentences.

A child who passes all the tests up to and including those for 6-year-olds and none beyond, has a "mental age" of 6, whatever his actual chronological age. Suppose, however, that he passes all the tests up to but not including the 6-year level, and then passes only three at the 5-year level and one at the 7-year level. His mental age is credited with 0.2 additional years for every item he passes beyond the level where he has passed them all. This child's mental age would be $5+0.6+0.2$, or 5.8 years of mental age. If his chronological age were 6 years, he would be slightly below average; if 5 years, somewhat above.

Binet did not come up with the "intelligence quotient" (IQ) itself; this fell to the German psychologist William Stern to do soon thereafter. Stern saw that a child who is 1 year behind at the age of 6 is more retarded than a child who is 1 year behind at the age of 13. It is the relation between mental and chronological age that matters, not just their difference, and this relation is best expressed by the ratio of the two numbers. To get the IQ, divide mental age by chronological age and multiply by 100 to get rid of the decimals. Thus, a 6-year-old child who comes through with a mental age of 9 is in these terms as bright as an 8-year-old with a mental age of 12, both having the impressive IQ of 150.

The IQ of 100 divides the population into two roughly equal groups. This is not a fact of nature but an outcome of how the tests were made. Binet and his successors picked and chose until they found items that the average child at each age could just pass, thus assuring that the average child's mental age equals his chronological age and his IQ 100. The idea of a mental age assumes that mental growth is accumulative and consecutive, so that a child who has mastered the items at a given age level one year will (barring disease or trauma) continue to do at least that well as he ages. In this case nature, not the testmakers, meets the condition. At each age during childhood we can do intellectually what we have done before, adding competence rather than replacing it.

Binet's idea for mental testing would not have worked for grubs and caterpillars, which appear to lose their grasp of burrowing and cocoon spinning as they become competent at flight. In other respects, too, Binet was fruitfully combining nature and artful design in his tests. Items on the test were included only if some children were ahead of their age in solving them, some behind, but the largest number were neither. Overall, the spread of performance conformed to the bell-shaped curve that statisticians call "normal," with about as many superior children as inferior, but with most crowding around the average.

Binet's ideas took hold powerfully and quickly. It was not only in France that the average 8-year-old child could just barely repeat accurately five digits read to him, for the Binet scale was readily exported to Belgium, Great Britain, American, Italy, and so on. The remarkable exportability of the tests was probably the first convincing argument for their soundness. Items that drew on bits of specific, seemingly arbitrary knowledge crossed national and linguistic boundaries as easily as the fundamental tests of memory and reasoning. It could be relied upon, for example, that the average 9-year-old would be able to name in order the months of the year. What does this say about the IQ? Would we downgrade a Papuan child, raised in New Guinea, if he could not name the months? Clearly not, if his language had no such names or had some different scheme for cutting up the year. Some of the items on a test are specific to a culture, but that does not make them poor items. A given test is only for people drawn from the same general population that the test was standardized on. Even if it is hard to locate the precise boundaries of this general population, a useful intelligence test should incorporate at least some of the material of a culture, or it may miss gaging the child's ability to assimilate his surroundings. Virtually every child grows up on some culture or another, and his intelligence score (if that concept is to retain its ordinary meaning) must reflect his sensitivity to it. The Papuan child cannot sensibly be tested on a western intelligence test. He would do poorly, but he would also do poorly in most other contacts with Western society. It would not mean that he was not intelligent. It would mean only that he was not meeting the underlying conditions of the test, which assume that he has been drawn from the standardizing population. Analogously, a child who gets a very high IQ after being drilled by parents or teachers on test items is probably not all that bright, and for the same reason. Like any other instrument of measurement, the IQ test must be used according to the directions. One does not use an oral thermometer after eating hot soup or sucking on ice cubes—not if one wants to know one's temperature. One may have a fever with a cool mouth, but the thermometer will not reveal it. So, the Papuan child may be bright or dull or average, but only a test standardized in his cultural environment can show which. It is not that "intelligence" itself is peculiarly European or North American, even if the instrument for gaging it is.

A person's IQ is a different sort of fact about him than his height or his weight or his speed in the hundred-yard dash, and not because of the difference between physical and mental attributes. Unlike inches, pounds, or seconds, the IQ is entirely a measure of relative

standing in a given group. No such relativism is tolerated for the conventional measures. Gulliver may have looked like a giant in Lilliput and a mite in Brobdingnag, but he was just about 70 inches tall wherever he went. Relativism is tolerated for the IQ because, first of all, we have nothing better. If the testers came up with something like a platinum yardstick for mental capacity, it would quickly displace the IQ. But more than this can be said for the IQ. Because the group with which a child is implicity compared is effectively the entire population of Western society, there is great stability to the comparison. The IQ gives one's standing among the people with whom one will live. And if it can be assumed that so large a sample of mankind is reasonably representative of the whole, then a relative measure is quite informative. An IQ of 100 would then indicate average intelligence, compared to people in general and not some small group; and IQ of 150 would denote high intelligence, and so on.

At around adolescence, people seem to stop acquiring new intellectual powers, as distinguished from new information or interests. For example, immediate memory span grows until the age of 15, but not thereafter. The average person can repeat seven digits at 15 or at 50. Other items in the Binet scale similarly level off at about the same age. Thus, if one were to continue calculating IQ in the same way, dividing a fixed mental age by a growing chronological age, one's score would plummet, reaching (for the average person) IQ 50 at about the age of thirty and IQ 25 at the age of 60 (assuming that the mental age is stuck at 15). To avoid such nonsense, some other measure of relative standing is often used for adults. Thus, instead of saying that a man has an IQ of 130, say instead that he tests higher than 96 percent of his peers, and then define the peer group. It can be all American adults, or Caucasions, or college graduates, or members of the United Auto Workers or the League of Women Voters. In fact, since the IQ is itself standardized on groups of peers (usually children), it and the percentile score are directly and simply translated one into the other.

Binet invented the modern intelligence test without saying what intelligence is. At first he was trying to sort out the mental defectives; later; he was trying to rate all the children—defective, average, or superior. Some rough-and-ready notion of intelligence lurked in the background—having to do with mental alertness, comprehension, speed, and so on—but he was not forced to defend an abstract definition in order to sell the idea of his test to the world. Instead, he could point to how well the test worked. Rarely did a bright child, as judged by the adults around him, score poorly, and rarely did a poor scorer seem otherwise bright. Occasionally a child would do worse than expected on the test because a teacher had confused obedience with brightness, or better than expected when rebelliousness had been mistaken for stupidity, but in general most children ended up about where they were expected to. The value of the test was that it gave an objective assessment about a child in an hour or so, and any trained technician could administer it. With the test as a yardstick, children who knew no one in common could be directly compared, for whatever purpose.

But is intelligence really an attribute, like height, that can be expressed in a single number? Even granting that IQ is a measure only

of relative standing, can relative standing be given in a single number? Is Jimmy really altogether brighter than Johnny if his IQ is higher? Perhaps Jimmy is brighter as regards A, B, C, and D, but Johnny has him beaten on E, F, and G. Even Binet admitted that intelligence was not just one thing; otherwise his labors in creating a test would have been far easier. Once, when he was speculating about the nature of intelligence, Binet mentioned the attributes of directedness, comprehension, inventiveness, and critical capacity, which he thought may vary somewhat independently from person to person. Usually, however, he was too busy with his practical goals to dwell on hypotheses.

Even as Binet was developing the first intelligence scale, others were grappling with the conceptually tougher problem of the structure of intelligence. The story of the key mathematical discoveries would be out of place here, but the highlights may be worth noting. An Englishman named Charles Spearman resigned a commission in the British Army after serving in the Boer War and set to work on the problem. Taking the intercorrelations between scores on simple mental tests as his basis, he concluded that there was a "universal" intellectual capacity—which he labelled "g" for "general"—plus a host of minor unrelated capacities of no great scope. The universal factor, he said, premeated all intellectual activity, while the others were variously absent or present in any given task. To be smart, for Spearman, mainly having lots of g. Although he had some evidence for this theory, it did not endure even for Spearman, who revised it after a decade or so. Nevertheless, his mathematical procedures were an essential link between Francis Galton's formulas for assessing correlation and the vastly more complex methods of "multiple factor analysis," which is the contemporary term.

Following Spearman, the next step was taken by L. L. Thurstone, an American electrical engineer who left a job in Edison's laboratory in East Orange, N.J., to work on psychological measurement. A long and illustrious career, covering the measurement not only of intelligence but also of attitudes, personality, sensory capacity, motivation, and the learning process was the result. For intelligence, Thurstone subdivided Spearman's general factor, g, into a set of primary mental abilities (PMA): spatial visualization, perceptual ability, verbal comprehension, numerical ability, memory, word fluency, and reasoning (inductive and deductive). These are just verbal labels tagged on at the end of a mathematical procedure that really has no verbal labels in it. It would be more precise (if less informative) to say that Thurstone found evidence for seven or eight separate factors as aspects of intelligence, and to leave it at that. With more powerful mathematics and more abundant data, Thurstone's successors have teased out new factors. Like nuclear physics with its proliferation of elementary particles, the study of intelligence has suffered from its riches. Now there are experts who find evidence of over 100 components in intelligence, and there is no sign of a limit.

Thurstone noted some intercorrelations among the primary mental abilities. People who excelled, for example, in verbal comprehension were often high in word fluency. Other constellations also kept turning up. Such correlations among the factors themselves could signify that mental abilities are hierarchical, arranged in layers. At the very

top, there may be a general intellectual power, like Spearman's g, pervading all mental activity. To be smart means having the power in abundance, to be stupid means having a shortage, so that all of Thurstone's PMA's will be to some degree correlated. At the next level down, the PMA's break into clusters involving either verbal abilities or numerical or logical abilities. Then there are the separate PMA's themselves, which vary somewhat independently despite their intercorrelations. In addition to being generally bright or stupid or average, people are verbal, numerical, imaginative, and so on. People can be so strong in one factor or another that they excel in some areas without any special abundance of g. And, inversely, some people may be so poorly endowed in one or the other factors that they appear occasionally incompetent, notwithstanding substantial g. Although the hierarchy seems like a plausible theory of intelligence, it will remain hypothetical until the experts agree on its specific features—which has yet to happen.

Even at best, however, data and analysis can take us only so far in saying what intelligence is. At some point, it becomes a matter of definition. For example, we would reject any intelligence test that discounted verbal ability or logical power, but how about athletic prowess or manual dexterity or the ability to carry a tune or qualities of heart and character? More data are not the final answer, for at bottom, subjective judgment must decide what we want the measure of intelligence to measure. So it is for all scales of measurement—physical as well as psychological. The idea of measuring length, weight, or time comes first; the instrument comes thereafter. And the instrument must satisfy common expectations as well as be reliable and practical. In the case of intelligence, common expectations center around the common purposes of intelligence testing—predicting success in school, suitability for various occupations, intellectual achievement in life. By this standard, the conventional IQ test does fairly well. The more complex measures, such as Thurstone's PMA's, add predictive power that is sometimes essential. As for what intelligence "really" is, the concept still has ragged edges where convenience and sheer intuition set boundaries that will no doubt change from time to time. The undisputed territory has, however, become formidable.

Most of us get our first, sometimes our only IQ test in school; the predictive power of the IQ is encountered first in our school grades; our teachers know our IQ's even when our parents (let alone we ourselves) do not. But for all these connections, IQ and education are only correlated, not identical. First of all, there is the fact of variability: at each level of education, the IQs span a broad range, and at each level of IQ among adults, the amount of education completed also spans a broad range. Moreover, school grades show the effect of the environment more than IQ's. And, finally, the correlation between IQ and schooling shows up even when the IQ is obtained from six-year-olds just starting school. Of course, once a child is known to have a high or low IQ, he may live up, or down, to his teachers' expectations, but even granting that complication, the IQ could hardly predict how much schooling there is going to be in someone's life if it were itself just a result of schooling.

The discrepancies between IQ and school grades are instructive, because they follow a definite pattern. It is not just that the IQ is not an

exact predictor of grades, but that children with low IQ's almost always do poorly in school, while children with high IQ's cover the range from excellent down to poor. For schoolwork, as for many other correlates of the IQ, intelligence is necessary but not sufficient. It is as if a high IQ offers merely the opportunity for scholastic achievement, but something more is needed to exploit it. We can guess what the something more might be—interest, emotional well-being, energy—but we do not know. Other activities that are correlated with IQ—such as success in business—also seem to call on something more, although perhaps not the same extras as good schoolwork does. No doubt it takes physical strength and stamina to be a champion athlete, but for many sports it takes some intelligence as well. To be a successful actor may take a good appearance or voice, but no doubt also intellect. The examples could be multiplied almost endlessly. IQ seems to be the sine qua non for an extraordinary variety of successes, but for virtually nothing practical is it the sole requirement.

But still, what is it? Even if it is not just schooling, may it not be a cryptic index of membership in the middle and upper classes, as many critics argue? We often hear that both IQ and successful education, and all the other correlates, follow from the more basic fact of social origin. To this criticism there is no short and simple answer. The correlation between IQ and social class (usually defined in terms of occupation, income, and patterns of personal association) is undeniable, substantial, and worth noting. A cautious conclusion, based on a survey of the scientific literature, is that the upper class scores about 30 IQ points above the lower class. A typical member of the upper class gets a score that certifies him as intellectually "superior," while a typical member of the lower class is a shade below average (that is, below IQ 100). Precise values cannot be taken too literally, for they depend on somewhat arbitrary definitions of social class and on which particular IQ test is used, but the basic finding is beyond dispute. Depending on whether one is for or against testing, one will see this class difference as a weakness either in the intellect of the underprivileged or in the tester's definition of intelligence. But in either case, there is no basis for assuming that no poor people have high IQ's. On the contrary, many members of the lower class must have superior IQ's, notwithstanding the low overall average. Recall, that, by design, there are as many people above IQ 100 as below. In contrast, the social scale is definitely lopsided, with many more at the bottom than at the top even in affluent America. Only about 10 percent of our people meet the criteria for the upper and upper-middle classes, while about 65 percent are in the working class and below, with the remainder in between. But only 50 percent of the people have subnormal (below 100) IQ's. And so, there must be at least 15 percent of our population in the bottom classes with supranormal (above 100) IQ's.

It is one thing to note the correlation between social class and IQ but something else to explain, or even interpret it. It does not prove that the IQ is caused by social class, any more than it proves the reverse—that social class is caused by IQ. More information is needed to sort out the possibilities. Since a family's social standing depends partly on the breadwinner's livelihood, there might be a further

correlation between IQ and occupation. A large sample of enlisted men in the Air Force in World War II, drawn from 74 different civilian occupations, revealed in detail the expected IQ differences. Here are some of the findings, culled from a study by T. W. Harrell and M. S. Harrell published in 1945 in a periodical called Educational and Psychological Measurement:

Rank in list of 74 occupations	*Civilian occupation*	*Average I.Q.*
1	Accountant	128.1
5	Auditor	125.9
10	Draftsman	122.0
15	Sales manager	119.0
20	Clerk-typist	116.8
25	Radio repairman	115.3
30	Laboratory assistant	113.4
35	Musician	110.9
40	Sales clerk	109.2
45	Power lineman	107.1
50	Riveter	104.1
55	Bartender	102.2
60	Molder	101.1
65	Baker	97.2
70	Lumberjack	94.7
74	Teamster	87.7

Each occupation has a range of IQ's: not-so-bright accountants and very bright bakers are far from unknown. But just as for good grades in school, a high IQ is necessary for some occupations, even if it is not sufficient. For example, among the 74 civilian occupations that turned up in the group, public relations proved to have the fourth highest average IQ, with the top IQ an impressive 149. The top truck driver also registered 149, but truck drivers averaged 67 in the list of occupations, close to the bottom with lumberjacks and teamsters. The lowest PR man had an IQ of 100, while the dullest truck driver tested an almost unbelievable 16—essentially no tested intelligence at all. So it was in general. The more prestigious occupations—law, engineering, science, public relations, and so on—seem to require a certain minimum IQ, well above the minimum for the less prestigious occupations—for the bakers, chauffeurs, barbers. As far as IQ alone is concerned, virtually anyone can be, for example, a welder, but half of mankind (the half below IQ 100) is not eligible for auditing, even if the brightest welder may equal the brightest auditor in IQ.

In this characteristic way, then, IQ affects one's occupation. And it is obvious that occupation affects one's social standing. It then follows logically that IQ affects social standing. When people are asked to rate the prestige of different occupations, they turn up with lists that look very much like the lists based on average IQ's—the professionals at the top, the laborers at the bottom, and the minor businessmen and

white-collar workers in the middle. These ratings have been as stable as the corresponding data on the IQ, in both America and Europe, and according to people up and down the social scale.

The ties among IQ, occupation, and social standing make practical sense. The intellectual demands of engineering, for example, exceed those of ditch digging. Hence, engineers are brighter, on the average. If virtually anyone is smart enough to be a ditch digger, and only half the people are smart enough to be engineers, then society is, in effect, husbanding its intellectual resources by holding engineers in greater esteem and paying them more. The critics of testing say that the correlations between IQ and social class show that the IQ test is contaminated by the arbitrary values of our culture, giving unfair advantage to those who hold them. But it is probably no mere coincidence that those values often put the bright people in the prestigious jobs. By doing so, society expresses its recognition, however imprecise, of the importance and scarcity of intellectual ability.

Binet's first scale served the humanitarian goal of getting retarded children into schools that would help them. But the test can also be used to spot exceptionally gifted children, for their own sake and society's. If the tests work, then gifted children should grow up to become unusually accomplished adults, just as the reverse is true at the other end of the scale.

The top of the scale provided the subject of a massive longitudinal study by Lewis M. Terman and his associates at Stanford University. For almost 40 years, they followed the lives of a large group of gifted people, publishing their results in five volumes between 1925 and 1959 under the general title of "Genetic Studies of Genius." The plan of the study was simple: find a large group of young children with exceptionally high IQ's, record as many potentially interesting and useful additional facts about them as practicable, and then follow the course of their lives. Terman and his staff found slightly more than 1,500 California children whose IQ's averaged about 150. (Because they used different intelligence scales for some of the children, no precise average figure can be given.) This was no small achievement in itself, for an IQ of 150 or greater is, a rarity, possessed, on the average, by the smartest child in a randomly selected group of about 200. Most of the children were between the ages of 8 and 12 when chosen, but there were also some younger anad some recruited in high schools.

Right from the start the findings were informative. For example, highly bright boys were easier to locate than highly bright girls. And the disparity increased slightly with age, suggesting that whatever the IQ is, boys maintain it better than girls. For this reason, the final sample had 857 boys and 671 girls. The children, mainly from urban public schools, definitely did not represent the ethnic or social composition of their communities. Compared to the population from which they were drawn, there was an enormous (over tenfold) excess of the children of fathers in the professions and an even more marked scarcity (only .013) of the children of laborers, echoing once again the correlation between IQ and social class. In addition, the sample contained an excess of Western and Northern Europeans and Jews, and a shortage of Latins, non-Jewish Eastern Europeans, and Negroes. Since the communities sampled had relatively few Orientals, it was

hard to tell whether too few, too many, or just the right number of gifted Oriental children turned up, statistically speaking.

The children were nonrepresentative physically as well as intellectually, ethnically, and socially. They tended to be taller, heavier, more broad-shouldered, stronger in hand grip, larger in the vital capacity of their lungs, and somewhat earlier in their sexual maturity than children in the general population. The physical differences, though not large, were large enough to counter the stereotype of the fragile bookworm. Not surprisingly, the gifted children did better in school than their classmates, but mainly in subjects—like reading and arithmetic—that seem to call on intelligence. In subjects like woodworking or sewing, the gifted children enjoyed no particular advantage. They most often liked precisely the subjects that the other children most often disliked, such as reading and arithmetic. At 7 years of age the gifted children were already reading books at a higher rate than the average child of 15. And even in sports they outdid their classmates, knowing more about the games of childhood and knowing about them earlier. Finally, even in tests of "character"—honesty, tendency toward overstatement, trustworthiness, and the like—the gifted children showed their precocity. At 9 or 10 years, they had reached the "moral development," by those no doubt quaint standards, of the average child of 13 or 14.

Children with IQ's of 150 or so are, then, special. But the big question is whether they mature into something special, for that would be the proper test of intelligence testing. Did the IQ make the difference it should have made? At last assessment, the sample had reached their middle forties, about 35 years after their selection for the study. The death rate in the sample had been less by a third than that in the general population, with fatal accidents quite uncommon. Childhood delinquency, criminal convictions, and alcoholism are all strikingly rare in the sample. More common, and benign, maladjustments are not so rare, with the women showing slightly more emotional trouble than the men. It may be a psychological burden to be so bright a woman in our culture, but this is pure speculation. In any event, not much can be made of the differences in minor mental disturbance between the sample and the general population.

About 70 percent of the sample finished college, men ahead of women by a couple of percentage points. This should be compared with the 8 percent of their contemporaries in the general population who finished college (the 1930–40 college generation). Out of the more than 1,500 in the sample, only 11 did not finish high school, and of these, eight went to professional or trade school. Forty percent of the male college graduates earned law, medical, or Ph. D. degrees, and over half of all the college graduates have at least some postgraduate training. There are, proportionately, five times as many Ph. D.'s in the sample as in the population of college graduates in general. As expected, the sample excelled in college: 80 percent averaging B or better in their courses, and more than 35 percent graduating with honors (Phi Beta Kappa, cum laude, or the like). In addition to their academic degrees, the sample has earned a disproportionately large number of professional licenses—CPA's, Fellows of the American Board of Surgery, Fellows of the American Institute of Architects, and so on.

The 10 most common occupations among the men are not the common lot in our society: lawyers first, followed by college faculty members, engineers, physicians, school administrators or teachers, chemists and physicists, authors, architects, geologists, and clergymen. All told, over 85 percent of the workingmen became either professionals or managers in business and industry, with the first category the larger. At the other end of the occupational scale, only about 3 percent became semiskilled laborers or farmers, and virtually none unskilled laborers. The men are bunched at the top of the scale of occupations, just as they are at the top of the scale of IQ. And the sample outperforms not only the population in general, but also the average college graduate. The run-of-the-mill college graduate has a 5 percent chance of becoming a semiskilled or unskilled laborer; the sample's college graduate has a chance of only 0.5 percent, a tenfold reduction.

Even with the relatively fewer employed women than men in the sample, the distinction of employment still shows. Approximately two thirds of the working women held professional positions—in universities, school and welfare systems, journalism, medical and paramedical professions, and so on. In addition, most of the women were married and raising children. The antifeminist threat that the education of women could remove the brightest potential mothers from the breeding stock receives no support in these results.

In addition to everything else, a high IQ pays in money. The average professional or managerial man in the sample was earning about $10,500 in 1954, compared to a national average of about $6,000 for those occupations. Even the semiskilled and clerical workers in the sample were outearning, by about 25 percent, the general averages for the same jobs. The total family income for the sample more than doubled that for white, urban American families of roughly the same socioeconomic status. About 30 percent of the families in the sample earned more than $15,000 a year in 1954, compared to only 1 percent for ordinary families in the same general socioeconomic class. The sample shows the economic advantages of a high IQ, after discounting education, race, occupation, and geography.

In the general population, income and education correlate highly. One encounters, from time to time, estimates of how much a high school diploma or a bachelor's degree should add to one's paycheck. No doubt about the facts—more highly educated people make more money—but the interpretation is arguable. The usual interpretation assumes that a given man with some higher education would earn more than the same man without the education. But that is not really what the data show, for we do not know whether the people who have the extra education are the same in other ways as those who do not. Suppose, for the sake of relevance, that income really depended more on IQ than on education. Suppose further that the amount of schooling also depended on IQ. Educated people would then earn more not because they were more educated (which they would also be) but because they were smarter (had a higher IQ). To disentangle the complex factors in society at large and find out what causes what, simple correlations are not enough. We need to know if income would be correlated with IQ if education were held constant, or conversely if income would be correlated with education if IQ were held constant. Terman's high

IQ sample is a step in this direction, for it allows us to see whether income depends on education when IQ is held constant at the virtual top of the scale.

High school graduates in the Terman sample were earning about as much as the college graduates with a bachelor's degree. Further schooling beyond the bachelor's did improve income somewhat. However, of the six men with the highest incomes (ranging upward of $100,000 a year) only one finished college. The highest annual income of all, $400,000 for the last year reported, was earned by a man who had had no college whatever. In other words, if very high income is your goal, and you have a high IQ, do not waste your time with formal education beyond high school. For this particular sample, education did not unequivocally add to income, as most often claimed. Perhaps for people of more ordinary talents the connection is straightforward, but the study shows that while IQ definitely affects income, education may not.

Not just the economic facts of life were gathered. When the men were asked about their state of mind, almost 90 percent said that they were at least fairly content, and virtually half were finding "deep satisfaction" in their lives. Only 6 percent reported discontentment. The more prosperous men were generally the more contended. When the men were asked to estimate how well they were living up to their intellectual abilities, there was again a correlation between satisfaction and income. The average yearly wage of those who said they were "fully" living up to their capacities was almost $12,000, while the group least satisfied was making less than $5,000 per year.

Women's salaries were substantially lower than the men's and did not correlate with contentment. Nothwithstanding their poorer salaries, on an average, the women reported greater satisfaction in their lives than the men. The housewives, who were earning less money than anyone else, expressed about as much satisfaction as any other group in the sample. There is little here to support the feminist argument that a housewife's life is intolerable, especially for educated, intelligent women. It would be hard to pick a brighter group than the women in this study, yet they seemed to be adjusting easily to their lot. To give but one example of the many striking cases, a woman whose IQ of 192 places her close to the top of the entire sample, and whose retested intelligence at maturity was again virtually at the top, was raising eight children, including three sets of twins. According to the account at the latest report, she had no outside activity at that time other than an interest in the PTA, but was apparently content, if not serene. Of course, such tranquillity may be gone now, 15 years later.

The enormous harvest from the sample in their middle forties included about 2,000 scientific and technical articles, 60 books, 33 novels, 375 short stories or plays, 325 miscellaneous publications, 230 patents, not to mention the hundreds of radio and television scripts, newspaper stories, pieces of art and music. Their names turn up disproportionately often in compilations of our effective people—in "Who's Who, American Men of Science," the National Academy of Sciences, and so on. They are active in P.T.A., clubs, hobbies. They vote far more faithfully (over 90 percent of the time in national elections) than the general population (and are somewhat more conservative). By the

mid-1950s, they had spawned (with spouses who where themselves significantly brighter than average) about 2,500 children whose average IQ appears to be above 130—not as brilliant as their exceptional parents, but still among the top 5 percent of the population. Even in their mid-forties, the sample continued to test within the top 1 percent of the general population in intelligence, whether or not they had been successful in their careers. No doubt the predictive power of the IQ is outlasting the first 35 years of the study.

No single study is beyond criticism, not even this massive enterprise by Terman and his associates. Critics can point to the possibility of hidden biases in the original selection of the children. Not every child in the California schools was tested, only those who looked "promising" for one reason or another. The final section employed just the IQ, but the prescreening may indeed have been a source of bias. Later estimates uncovered, however, only a few children missed this way, certainly not enough to change the general conclusions about the predictiveness of IQ. Critics may also wonder how people are affected by being included in this select group. Are they impelled to excel, or are they stunted by anxiety? Judging from all the other data showing correlations between IQ and achievement, the sample seems to be about normal for an IQ of 150. Whatever the flaws in the study, there can be no reasonable doubt about its main conclusion. An IQ test can be given in an hour or two to a child, and from this infinitesimally small sample of his output, deeply important predictions follow—about schoolwork, occupation, income, satisfaction with life, and even life expectancy. The predictions are not perfect, for other factors always enter in, but no other single factor matters as much in as many spheres of life.

Terman was unapologetic about where he thought IQ comes from. He believed in the inheritance of IQ, at least to a considerable degree. Bluntly, but not dogmatically, he wrote in 1925:

> There are . . . many persons who believe that intelligence quotients can be manufactured to order by the application of suitable methods of training. There are even prominent educators and psychologists who are inclined to regard such a pedagogical feat as within the realm of possibility, and no one knows that it is not. If it is possible it is time we were finding out. Conclusive evidence as to the extent to which IQ's can be artificially raised could be supplied in a few years by an experiment which would cost a few hundred thousand or at most a few million dollars. The knowledge would probably be worth to humanity a thousand times that amount.

The opening paragraphs of the disturbing and controversial article by Prof. Arthur R. Jensen of the University of California, Berkeley, could be taken as the equally blunt answer to Terman's challenge, 44 years later.

> Compensatory education has been tried and it apparently has failed.
>
> Compensatory education has been practiced on a massive scale for several years in many cities across the Nation. It began with auspicious enthusiasm and high hopes of edu-

> cators. It had unprecedented support from Federal funds. It had theoretical sanction from social scientists espousing the major underpinning of its rationale: the "deprivation hypothesis," according to which academic lag is mainly the result of social, economic, and educational deprivation and discrimination—an hypothesis that has met with wide, uncritical acceptance in the atmosphere of society's growing concern about the plight of minority groups and the economically disadvantaged.
>
> The chief goal of compensatory education—to remedy the educational lag of disadvantaged children and thereby narrow the achievement gap between "minority" and "majority" pupils—has been utterly unrealized in any of the large compensatory education programs that have been evaluated so far.

And the reason, Jensen goes on to say, why compensatory education has failed is that it has tried to raise IQ's which, he argues, are more a matter of inheritance than environment, and therefore not very amenable to corrective training. What evidence has he for this unexpected and unpopular conclusion?

The problem with nature and nurture is to decide which—inheritance or environment—is primary, for the IQ is exclusively the result of neither one alone. Advocates of environment—the clear majority of those who express themselves publicly on the subject—must explain why IQ's usually stay about the same during most people's lives and also why high or low IQ's tend to run in families. Those facts could easily be construed as signs of a genetic basis for the IQ. The usual environmentalist answer argues that IQ's remain the same to the extent that environments remain the same. If you are lucky enough to be well-born, then your IQ will show the benefits of nurturing, which, in turn, gives you an advantage in the competition for success. If, on the other hand, you are blighted with poor surroundings, your mental growth will be stunted and you are likely to be stuck at the bottom of the social ladder. By this view, parents bequeath to their children not so much the genes for intelligence as the environment that will promote or retard it.

In one plausible stroke the environmetalist arguments seem to explain, therefore, not only the stability of the IQ but also the similarity between parents and children. The case is further strengthened by arguing that early training fixes the IQ more firmly than anything we know how to do later. And then to cap it off, the environmentalist may claim that the arbitrary social barriers in our society trap the underprivileged in their surroundings while guarding the overprivileged in theirs. Anyone who accepts this series of arguments is unshaken by Jensen's reminder that compensatory education has failed in the United States, for the answer seems to be ready and waiting. To someone who believes in the environmental theory, the failure of compensatory education is not disproof of his theory, but rather a sign that we need more and better special training earlier in a person's life.

To be sure, it seems obvious that poor and unattractive surroundings will stunt a child's mental growth. To question it seems callous. But even if it is plausible, how do we know it is true? By what evi-

dence do we test the environmentalist doctrine? The simplest possible assessment of the inherited factor in IQ is with identical twins, for only environmental differences can turn up between people with identical genes. In an article recently published in the periodical "Behavior Genetics," Professor Jensen surveys four major studies of identical twins who were reared in separate homes. Most of the twins had been separated by the age of 6 months, and almost all by the age of 2 years. The twins were Caucasians, living in England, Denmark, and the United States—all told, 122 pairs of them. The overall IQ of of the 244 individuals was about 97, slightly lower than the standard 100. Identical twins tend to have slightly depressed IQ's, perhaps owing to the prenatal hazards of twindom. The 244 individuals spanned the range of IQ's from 63 to 132, a range that brackets most of humanity—or to be more precise, 97 percent of the general population on whom intelligence tests have been standardized.

Being identical twins, the pairs shared identical genetic endowments, but their environments could have been as different as those of random pairs of children in the society at large. Nevertheless, their IQ's correlated by about 85 percent, which is more than usual between ordinary siblings or even fraternal twins growing up together with their own families. It is, in fact, almost as big as the correlations between the heights and weights of these twins, which were 94 percent and 88 percent respectively. Even environmentalists would expect separately raised twins to look alike, but these results show that the IQ's match almost as well. Of course if the environment alone set the IQ, the correlations should have been much smaller than 85 percent. It would, however, be rash to leap to the conclusion that the 85 percent correlation is purely genetic, for when twins are placed into separate homes, they might well be placed into similar environments. The children had been separated not for the edification of psychologists studying the IQ, but for the weighty reasons that break families up—illness, poverty, death, parental incapacity, and so on—and the accidents of separation may not have yielded well-designed experiments. Some of the pairs were no doubt raised by different branches of the same family, perhaps assuring them considerable environmental similarity anyway. In such cases, the correlation of 85 percent would not be purely genetic, but at least partly environmental. Fortunately for our state of knowledge, one of the four studies examined by Jensen included ratings of the foster homes in terms of the breadwinner's occupation. Six categories sufficed: higher professional, lower professional, clerical, skilled, semiskilled, unskilled. Now, with this classification of homes, we know a little about whether the twins were raised in homes with a similar cultural ambience. To the extent that the environment in a home reflects the breadwinner's occupation, the answer is unequivocally negative, for there was literally no general correlation in the occupational levels of the homes into which the pairs were separated. At least for this one study—which happened to be the largest of the four—the high correlation in IQ resulted from something besides a social-class correlation in the foster homes, most likely the shared inheritance.

Twins raised apart differ on the average by about seven points in IQ. Two people chosen at random from the general population differ by seventeen points. Only four of the 122 pairs of twins differed by as

much as 17 points. Ordinary siblings raised in the same household differ by 12 points. Only 19 of the 122 twin pairs differed by as much as that. And finally, fraternal twins raised in the same home differ by an average of 11 points, which was equaled or exceeded by only 23 of the 122 pairs. In other words, more than four times out of five the difference between identical twins raised apart fell short of the average difference between fraternal twins raised together by their own parents. At the same time, those separated twins were not so similar in schoolwork. Identical twins raised together resemble each other in both IQ and school grades. When twins are separated, their IQ's remain quite close, but their grades diverge. It seems that school performance responds to the environment substantially more than does the IQ, although neither one is solely the outcome of either nature or nurture.

The comparison between IQ and grades was one theme of Jensen's controversial earlier article. "How Much Can We Boost IQ and Scholastic Achievement?", which appeared in the winter of 1969 in the Harvard Educational Review. Jensen answered the title's rhetorical question about IQ with a scholarly and circumspect form of "not very much." The article is cautious and detailed, far from extreme in position or tone. Not only its facts but even most of its conclusions are familiar to experts. The failure of compensatory education was the occasion for the article, which served especially well in assembling many scattered but pertinent items. Jensen echoes most experts on the subject of the IQ by concluding that substantially more can be ascribed to inheritance than environment. Since the importance of inheritance seems to say something about racial differences in IQ that most well-disposed people do not want to hear, it has been argued that Jensen should not have written on the subject at all or that the Harvard Educational Review should not have, as it did, invited him to write on it.

Some of Jensen's critics have argued that because environment and inheritance are intertwined, it is impossible to tease them apart. The criticism may seem persuasive to laymen, for nature and nurture are indeed intertwined, and in just the way that makes teasing them apart most difficult. For intelligence—unlike, for example, skin color—the main agents of both nature and nurture are likely to be one's parents. One inherits skin color from one's parents, but the relevant environment does not come directly from them but from sun, wind, age, and so on. For skin color, resemblance to parents signifies (albeit not infallibly) inheritance; for intelligence, resemblance is ambiguous. Nevertheless analysis is possible even with IQ, as Jensen and his predecessors have shown. The most useful data for the purpose are the correlations between IQ and kinship, as exemplified by the twin studies, which set genetic similarity high and environmental similarity low. Foster children in the same home define the other extreme of kinship and environment. If environment had no bearing at all on intelligence, then the IQ's of such unrelated children should correlate slightly at most (and only to the extent caused by a special factor to be mentioned shortly). In contrast, if environment were all, then the correlation should approach the value for natural siblings. Actually, the IQ's of foster children in the same home correlate by about 24 percent (less than half the value for natural siblings). However, even the correlation of 24 percent cannot be credited entirely to the children's shared en-

vironment. Bear in mind that adoption agencies try to place "comparable" children in the same home, which means that there is more than just their common surrounding making them alike. Suppose, for example, that adoption agencies tried to put children with similar hair color in any given family. They could check on the natural parents, and perhaps even on the grandparents, and make a reasonable guess about the baby's eventual hair color. The foster children in a given home would then often have similar hair color; they would be unrelated by blood, but the similarity would be more genetic than environmental. By trying for a congenial match between foster child and foster parents—in appearance and in mental ability—adoption agencies make the role of environment look more important that it probably is.

In between foster siblings and identical twins come the more familiar relations, and these too have been scrutinized. If intelligence were purely genetic, the IQ's of second cousins would correlate by 14 percent and that of first cousins by 18 percent (the reasons for those peculiar percentages are well beyond the scope of this article, so they are offered without proof). Instead of 14 percent and 18 percent, the actual correlations are 16 percent and 26 percent—too large for genetic influences alone, but in the right range. Uncle's (or aunt's) IQ should, by the genes alone, correlate with nephew's (or niece's) by a value of 31 percent; the actual value is 34 percent. The correlation between grandparent and grandchild should, on genetic grounds alone, also be 31 percent, whereas the actual correlation is 27 percent, again a small discrepancy. And finally for this brief survey, the predicted correlation between parent and child, by genes alone, is 49 percent, whereas the actual correlation is 50 percent using the parents' adult IQ's and 56 percent using the parents' childhood IQ's—in either case too small a difference to quibble about. Parents and their children correlate about as well whether the children are raised at home or by a foster family, which underscores the relative unimportance of the environment.

The foregoing figures are lifted directly out of Jensen's famous article, figures that he himself culled from the literature of intelligence testing. The measurements say that (1) the more closely related by blood two people are, the greater the correlation between their IQ's and (2) the correlations fall in the right range from the purely genetic standpoint. By evaluating the total evidence, and by a procedure too technical to explain here, Jensen concluded (as have most of the other experts in the field) that the genetic factor is worth about 80 percent and that only 20 percent is left to everything else—the social, cultural, and physical environment, plus illness, prenatal factors, and what have you.

Jensen's two papers leave little doubt about the heritability of IQ among North American and Western European whites, whom most data on the subject describe. In fact, there is little dispute on this score, even among those who object vigorously to this work. It is the relation between heritability and racial differences that raises the hackel. Given the well-established, roughly 15-point black-white difference in IQ, the argument is whether the difference arises in the environment or the genes. If intelligence were entirely genetic, then racial differences would be genetic simply because they could be due to nothing else. Con-

versely, if the genes were irrelevant, then the racial difference would have to be due to the environment, again because there would be no alternative. As it is, IQ reflects both a person's genes and his environment. The racial issue really poses the nature-nurture question all over again, but this time for a particular finding—the higher scores of whites over black on IQ tests.

In general—not just the racial issue—the question of nature and nurture boils down to the study of variation. Granted that IQ's vary among people, to what extent does the variation correlate with the differences in their surroundings on the one hand and with the differences in their genetic makeup on the other? No one disputes the existence of all three kinds of variation—in IQ environment, and inheritance—only their interconnections. In effect, the environmentalist is saying that among a group of people, the various IQ's reflect the various surroundings more or less without regard to the genes. In contrast, the nativist is saying the reverse—that different IQ's reflect different genetic endowments rather than different environments. The study of quantitative genetics contrives to answer such riddles, and so a brief didactic excursion is in order. But instead of starting the lesson with IQ, let us consider a trait which we are not emotionally committed to begin with.

Suppose we wanted to know the heritability of skin color. We would not need science to tell us that dark or fair complexions run in certain families or larger groups. Nor must we be told that nongenetic elements also enter in, as when a person is tan the sun or pale with illness or yellow from jaundice or red with rage or blue with cold. The task of quantitative genetics is to come up with a number that says how large a role inheritance plays in the total amounts of variation in skin color that we see in a particular group of people at a particular time. If the number is large, then skin color is largely heritable; if very small, then the heritability is negligible. If the number is large, then there will be marked family resemblances; if small, then members of given families will be no more alike than unrelated people. To convey such information, the number must reflect which group of people we choose to study. Consider first the United States, with its racial and ethnic diversity. Much skin variation here is related to ancestry, whether black, white, yellow, red, or Mediterranean, Nordic, Alpine, or some blend. Family resemblances in skin color are quite strong in America, so the heritability should come out large. Now contrast this with an isolated village in Norway, full of Scandanavians with generations of pale-skinned ancestors. In the Norwegian town, whatever little variation there is in skin color is likely to be environmental, due to the circumstances of life rather than to the accident of inheritance. As regards skin color, children will be no more like their parents than their nonrelatives, so heritability should come out low.

The hardest thing to grasp about heritability is that it says something about a trait in a population as a whole, not about the relation betw n particular parents and their offspring. Skin color turns out to be more heritable in the United States than in Norway, even though the physiological mechanisms of inheritance are surely the same. In the Norwegian town, a swarthy father and mother (who probably got that way from exposure to the weather) are likely to have children as fair-skinned as their neighbors. In the American town, however, it is

more likely that the swarthiness of swarthy parents is genetic and will be passed on to the children. Although heritability is not the strictly physiological concept that laymen imagine it to be, it is uniquely useful for talking about the nature-nurture question, for it tells us whether traits run in families within a broader population of individuals.

The technical measure of heritability is a number between 0 and 1 that states how much of the variation in a trait is due to genetic factors. How it is calculated need not detain us here. It is enough to know that a heritability of 0.5 means (omitting some technical complexities) that the variation is due half to genetic factors and half to other factors; a heritability of 0.2 means that only a fifth of the variation is genetic, and so on. Some actual heritabilities of traits in animals may be helpful. In piebald Holstein cattle, for example, the amount of white in the fur has a heritability of about 0.95, a value so high that it is almost right to say the environment plays no role here. In contrast, milk yield has a heritability of only 0.3. White in the fur, therefore, breeds more true than milk production. In pigs, the thickness of body fat has a heritability of 0.55, while the liter size has a heritability of only 0.15.

Now back to IQ and the racial issue. Using the procedures of quantitative genetics, Jensen (and most other experts) estimates that IQ has a heritability between 0.80 and 0.85, but this is based almost entirely on data from whites. We may, therefore, say that 80 to 85 percent of the variation in IQ among whites is due to the genes. Because we do not know the heritability for IQ among blacks, we cannot make a comparable statement for them. But let us simply assume, for the sake of discussion, that 0.8 is the heritability for whites and blacks taken together. What could we say about the racial difference in IQ then? The answer is that we could still say nothing positive about it. Recall that the concept of heritability applies to a population as a whole. All we could say is that the differences between people, on the average and without regard to color, are 80 percent inherited. But within this broad generality, particular differences could and would be more or less inherited. Take, for example, the differences in IQ between identical twins. Even with the average heritability equal to 0.8, all twin differences have to be totally environmental, since their genes cannot differ. Or conversely, consider the differences between foster children in a given foster family. Because they are growing up in the same home, their IQ differences could easily be relatively more genetic than those of people taken at random. When this line of reasoning is applied to a racial (or ethnic) difference in IQ, the only proper conclusion is that we do not know whether it is more genetic, less genetic, or precisely as genetic as implied by a heritability of 0.8.

Jensen notes that we lack a good estimate of the heritability of intelligence among blacks. Although there are scraps of evidence for a genetic component in the black-white difference, the overwhelming case is for believing that American blacks have been at an environmental disadvantage. To the extent that variations in the American social environment can promote or retard IQ, blacks have probably been held back. But a neutral commentator (a rarity these days) would have to say that the case is simply not settled, given our present stage of knowledge. To advance this knowledge would not be easy, but it could certainly be done with sufficient ingenuity and hard work. To anyone

who is curious about this question and who feels competent to try to answer it, it is at least irritating to be told that the answer is either unknowable or better not known, and both enjoinders are often heard. And there is, of course, a still more fundamental issue at stake, which should concern even those who are neither curious about nor competent to study racial differences in IQ. It is whether inquiry shall (again) be shut off because someone thinks society is best left in ignorance.

Setting aside the racial issue, the conclusion about intelligence is that, like other important though not necessarily vital traits, it is highly heritable. It is not vital in the sense that it may vary broadly without markedly affecting survival, although it no doubt affects one's life-style. Does it do us any practical good to know how heritable intelligence is? We are not, for example, on the verge of Galton's vision of eugenics, even though we now have the mental test that he thought was the crucial prerequisite. For good or ill, and for some time to come, we are stuck with mating patterns as people determine them for themselves. No sensible person would want to entrust State-run human breeding to those who control today's States. There are, however, practical corollaries of this knowledge, more humble than eugenics, but ever more salient as the growing complexity of human society makes acute the shortage of high-grade intellect.

Heritability is first and foremost the measure of breeding true, useful for predicting how much of some trait the average offspring in a given family will have. For example, to predict the IQ of the average offspring in a family:

1. Average the parents' IQ's.
2. Subtract 100 from the result.
3. Multiply the result of (2) by 0.8 (the heritability).
4. Add the result of (3) to 100.

Thus, given a mother and father each with IQ's of 120, their average child will have an IQ of 116. Some of their children will be brighter and some duller, but the larger the family, the more nearly will the average converge into 116. With parents averaging an IQ of 80, the average child will have an IQ of 84. The formula predicts something that the experts call "regression toward the mean," the tendency for children to be closer to the general population average (in this case, IQ 100) than their parents. And in fact, very bright parents have children who tend to be merely bright, while very dull parents tend to have them merely dull. The amount of regression for a trait depends on the heritability—with high heritability, the regression is smaller than with low. Also, for a given trait the regression is greater at the extremes of a population than at its center. In other words, ordinary parents are more like their children (on the average) than extraordinary ones (whether extraordinarily high or low). All of these characteristics of the "generation gap" follow directly and completely from the simple formula given above. Thus, when the parents average 120, the regression effect is only four IQ points, but if they averaged 150, the regression effect would be 10 points. In comparison, height, with its herritability of 0.95, would show smaller regression effects than IQ, since the multiplier in step 3 of the formula is closer to 1. But even so, very tall parents tend to have children who are merely tall, and

very short parents tend to have them merely short. As long as the heritability of a trait falls short of 1, there is some regression effect.

Intelligence may be drifting up or down for environmental reasons from generation to generation, notwithstanding the high heritability. Height, for example, is said to be increasing—presumably because of diet and medicine—even with its 0.95 heritability. We can easily tell whether there has been a change in height, for the measures are absolute, and there is the tangible evidence of clothing, furniture, coffins, and the skeletons themselves. For intelligence, however, we have no absolute scales, only relative ones, and the tangible remains of intelligence defy interpretation. But if height has changed, why not intelligence? After all, one could argue, the IQ has a heritability of only 0.8, measurably lower than that of height, so it should be even more amenable to the influence of the environment. That, to be sure, is correct in principle, but the practical problem is to find the right things in the environment to change—the things that will nourish the intellect as well as diet does height. The usual assumption, that education and culture are crucial, is running into evidence that the physical environment—for example, early diet—might be more important. In fact, the twin studies that Jensen surveyed showed that the single most important environmental influence on IQ was not education or social environment, but something prenatal, as shown by the fact that the twin heavier at birth usually grew up with the higher IQ.

Suppose we do find an environmental handle on IQ—something, let us say, in the gestating mother's diet. What then? Presumably society would try to give everyone access to the favorable factor, within the limits of its resources. Intelligence would increase according. But that would not end our troubles with IQ. Recall that heritability is a measure of relative variation. Right now, about 80 percent of the variation in IQ derives from the genes. If we make the relevant environment much more uniform (by making it as good as we can for everyone), then an even larger proportion of the variation in IQ will be attributable to the genes. The average person would be smarter, but intelligence would run in families even more obviously and with less regression toward the mean than we see today. It is likely that the mere fact of heritability in IQ is socially and politically important, and the more so the higher the heritability.

The specter of communism was haunting Europe, said Karl Marx and Friedrich Engels in 1848. They could point to the rise of egalitarianism for proof. From Jefferson's "self-evident truth" of man's equality, to France's "*egalite*" and beyond that to the revolutions that swept Europe as Marx and Engels were proclaming their manifesto, the central political fact of their times, and ours, has been the rejection of aristocracies and privileged classes, or special rights for "special" people. The vision of a classless society was the keystone of the Declaration of Independence as well as the Communist Manifesto, however different the plan for achieving it.

Against this background, the main significance of intelligence testing is what it says about a society built around human inequalities. The message is so clear that it can be made in the form of a syllogism:

1. If differences in mental abilities are inherited, and
2. If success requires those abilities, and
3. If earnings and prestige depend on success,

4. Then social standing (which reflects earnings and prestige) will be based to some extent on inherited differences among people.

The syllogism has five corollaries, which make it more relevant to the future than to the past or present.

(*a*) As the environment becomes more favorable for the development of intelligence, its heritability will increase, as the preceding section showed. Regardless of whether this is done by improving educational methods, diet for pregnant women, or whatever, the more advantageous we make the circumstances of life, the more certainly will intellectual differences be inherited. And the greater the heritability, the greater the force of the syllogism.

(*b*) All modern political credos preach social mobility. The good society should, we believe, allow people to rise (and, by implication if not by frank admission, fall) according to their own efforts. The social barriers of the past—race, religion, nationality, title, inherited wealth—are under continuous assault, at least in principle. The separation of church and State, the graduated income tax, the confiscatory inheritance tax, the laws against discrimination and segregation, the abolition of legal class and caste systems all manifest a desire to accelerate movement on the social ladder. The standard wisdom of our time avows that people should be free of "unfair" impediments and divested of "unfair" advantages in all their endeavors. But the syllogism becomes more potent in proportion to the opportunities for social mobility, for it is only when able and energetic individuals can rise and displace the dull and sluggish ones that there can be sorting out of people according to inherited differences. Actual social mobility is blocked by innate human differences after the social and legal impediments are removed.

(*c*) It was noted earlier that there are many bright but poor people even in affluent America. The social ladder is tapered steeply, with far less room at the top than at the bottom. The obvious way to rescue the people at the bottom is to take the taper out of the ladder, which is to say, to increase the aggregate wealth of society so that there is more room at the top. This is, of course, just what has been happening since the Industrial Revolution. But one rarely noted byproduct of poverty is that it minimizes the inherited differences between classes by assuring that some bright people will remain at the bottom of the ladder. As the syllogism implies, when a country gains new wealth, it will tend to be gathered in the hands of the natively endowed. In other words, the growth of wealth will recruit for the upper classes precisely those from the lower classes who have the edge in native ability. Whatever else this accomplishes, it will also increase the IQ gap between upper and lower classes, making the social ladder even steeper for those left at the bottom.

(*d*) Technological advance changes the marketplace for IQ. Even if every single job lost in automating a factory is replaced by a new job someplace else in a new technology, it is more than likely that some of those put out of the old jobs will not have the IQ for the new ones. Technological unemployment is not just a matter of "dislocation" or "retraining" if the jobs created are beyond the native capacity of the newly unemployed. It is much easier to replace men's muscles with machines than to replace their intellects. The computer visionaries believe that their machines will soon be doing our thinking

for us too, but in the meantime, backhoes are putting ditchdiggers out of work. And the ones who stay out of work are most likely the ones with the low IQ's. The syllogism implies that in times to come, as technology advances, the tendency to be unemployed may run in the genes of a family about as certainly as bad teeth do now.

(*e*) The syllogism deals manifestly with intelligence. The invention of the intelligence test made it possible to gather the data necessary to back up the three premises. However, there may be other inherited traits that differ among people and contribute to their success in life. Such qualities as temperament, personality, appearance, perhaps even physical strength or endurance, may enter into our strivings for achievement and are to varying degrees inherited. The meritocracy concerns not just inherited intelligence, but all inherited traits affecting success, whether or not we know of their importance or have tests to gage them.

The syllogism and its corollaries point to a future in which social classes not only continue but become ever more solidly built on inborn differences. As the wealth and complexity of human society grow, there will be precipitated out of the mass of humanity a low-capacity (intellectual and otherwise) residue, that may be unable to master the common occupations, cannot compete for success and achievement, and are most likely to be born to parents who have similarly failed. In Aldous Huxley's "Brave New World," it was malevolent or misguided science that created the "alphas," "gammas," and the other distinct types of people. But nature itself is more likely to do the job or something similar, as the less well-known but far more prescient book by Michael Young, "The Rise of the Meritocracy," has depicted. Young's social-science-fiction tale of the antimeritocratic upheavals of the early 21st century is the perfect setting for his timely neologism, the word "meritocracy." The troubles he anticipated, and that the syllogism explains, have already caught the attention of alert social scientists, like Edward Banfield, whose book "The Unheavenly City" describes the increasingly chronic lower class in America's central cities. While Sunday supplements and popular magazines crank out horror stories about genetic engineering (often with anxious but self-serving testimonials from geneticists), our society may be sorting itself willy-nilly into inherited castes. What is most troubling about this prospect is that the growth of a virtually hereditary meritocracy will arise out of the successful realization of contemporary political and social goals. The more we succeed in achieving relatively unimpeded social mobility, adequate wealth, the end of drudgery, and wholesome environment, the more forcefully does the syllogism apply.

Are there alternatives short of turning back to social rigidity, poverty, drudgery, and squalor? The first two premises of the syllogism cannot sensibly be challanged, for they are true to some extent now and are likely to become more so in time. The heritability of intelligence will grow as the conditions of life are made more uniformly wholesome; intelligence will play an increasingly important role in occupational success as the menial jobs are taken over by machines. The one even plausible hope is to block the third premise by preventing earnings and prestige from depending upon successful achievement. The socialist dictum, "From each according to his ability, to each accord-

ing to his needs," can be seen as a bald denial of the third premise. It states that, whatever a person's achievement, his income (economic, social, and political) is unaffected by his success. Instead, the dictum implies, people will get what they need however they perform, but only so long as they fulfill their abilities. Those in power soon discover that they must insist on a certain level of performance, for what the dictum neglects is that "ability" is, first of all, widely and innately variable, and secondly, that it expresses itself in labor only for gain. In capitalist countries, the gain is typically in material wealth, but even where the dictum rules (if such places exist), social and political influence or relief from threat would be the reward for accomplishment. Human society has yet to find a working alternative to the carrot and the stick. Meanwhile, the third premise assures the formation of social classes.

Classlessness is elusive because people vary and because they compete for gain—economic and otherwise. The tendency to respect, honor, remunerate, and perhaps even envy people who succeed is not only ingrained but is itself a source of social pressure to contribute to one's limit. Imagine, for example, what would happen if the gradient of gain were inverted by government fiat. Suppose bakers and lumberjacks got the top salaries and the top social approval, while engineers, physicians, lawyers, and business executives got the bottom. Soon thereafter, the scale of IQ's would also invert, with the competition for the newly desirable jobs now including people with the highest IQ's. (For simplicity's sake, only IQ is mentioned, but there may be and no doubt are, other factors that contribute to success, for recall that IQ is only necessary, not sufficient.) The top IQ's would once again capture the top of the social ladder. But no government (let alone people themselves) is likely to conduct such an experiment, for it is not a sensible allocation of a scarce resource like high-grade intelligence. Nor could a government long equalize the gains from all occupations. It was noted before that the premium given to lawyers, doctors, engineers, and business managers is not accidental, for those jobs are left to incompetents at our collective peril. There are simply fewer potentially competent physicians than barbers. The gradient of occupations is, then, a natural measure of value and scarcity. And beneath this gradient is a scale of inborn ability, which is what gives the syllogism its unique potency.

It seems that we are indeed stuck with the conclusion of the syllogism. The data on IQ and social-class differences show that we have been living with an inherited stratification of our society for some time. The signs point to more rather than less of it in the future, assuming that we are not plunged back into a state of primeval poverty by some cataclysm or do not turn back to rigidity and arbitrarily privileged classes. Recall that regression toward the mean depends upon the heritability and that improving the environment raises the heritability. The higher the heritability, the closer will human society approach a virtual caste system, with families sustaining their position on the social ladder from generation to generation as parents and children are more nearly alike in their essential features. The opportunity for social mobility across classes assures the biological distinctiveness of each class, for the unusual offpring—whether more or less able than his or her) closest relatives—would quickly rise above his family or

sink below it, and take his place, both biologically and socially, with his peers.

If this is a fair picture of the future, then we should be preparing ourselves for it instead of railing against its dawning signs. Greater wealth, health, freedom, fairness, and educational opportunity are not going to give us the egalitarian society of our philosophical heritage. It will instead give us a society sharply graduated, with ever greater innate separation between the top and the bottom, and ever more uniformity within families as far as inherited abilities are concerned. Naturally, we find this vista appalling, for we have been raised to think of social equality as our goal. The vista reminds us of the world we had hoped to leave behind—aristocracies, privileged classes, unfair advantages and disadvantages of birth. But it is different, for the privileged classes of the past were probably not much superior biologically to the downtrodden, which is why revolutions had a fair chance of success. By removing arbitrary barriers between classes, society has encouraged the creation of biological barriers. When people can freely take their natural level in society, the upper classes will, virtually by definition, have greater capacity than the lower.

The measurement of intelligence is one of the yardsticks by which we may assess the growing meritocracy, but others tests of human potential and performance should supplement the IQ in describing a person's talents, interests, skills, and shortcomings. The biological stratification of society would surely go on whether we had tests to gage it or not, but with them a more humane and tolerant grasp of human differences is possible. And at the moment, that seems our best hope.

ASSESSMENT OF RACIAL DESEGREGATION IN THE BERKELEY SCHOOLS

By Arthur R. Jensen
University of California, Berkeley

BACKGROUND

In 1967, the Berkeley Board of Education voted to racially desegregate all of its elementary schools (grades K through six). Berkeley has had only one senior high school (grades 10 through 12) which had always been completely racially integrated. The three junior high schools (grades seven through nine) had been desegregated in 1964 by making one of the three junior highs for ninth grade only and adjusting school boundaries so that the remaining two schools would have similar racial compositions. The methods for desegregating the elementary schools evolved during the first half of 1967. They were approved by the board in January 1968 and were put into effect the following September. The desegregation plan consisted of two-way busing within broad attendance zones that geographically cut across the racial and socioeconomic stratification of the community. The largest school in each zone enrolled all children in grades four to six, while the smaller schools accommodated grades K to three. Some 3,500 of Berkeley's 9,000 elementary school children are bused each day. De facto segregation within schools is prevented by the district's explicit and enforced policy that all classes "will be heterogeneous by race, sex, academic performance, and when possible, socioeconomic level."

PLANS FOR EVALUATION

In fall 1967, while plans were evolving for desegregating the elementary schools, a number of citizens and groups appeared at the open meetings of the Berkeley Board of Education to urge that a part of the funds appropriated for accomplishing desegregation be used to conduct a thorough study of the effects of desegregation on the scholastic performance of minority and majority children during the first few years of the program. The board was urged to consider the merits of an impartial, independently conducted evaluation of the new program. Funds for the program from Federal and State agencies were also contingent, in part, upon the schools' having some plans for evaluating the programs which the funds were intended to support.

In pursuing means for setting up an independent evaluation, some members of the board and school officials approached Dr. Roger Heyns, then chancellor of the Berkeley campus of the University of California, to solicit the university's cooperation in formulating and carrying out an evaluation study. Chancellor Heyns formed an ad-

visory committee made up of several deans and directors of professional schools and research institutes of the university which had obvious interests in schools and community affairs, such as the schools of education, criminology, and social welfare, and the survey research center. The advisory committee held a number of meetings in fall 1967, to consider the feasibility of the university's undertaking the large-scale evaluation required by the Berkeley schools, and they conferred with a number of the faculty in relevant departments about the strategy and probable costs and personnel requirements of such a project. I was first called to confer with Chancellor Heyns and the advisory committee in November 1967. As a professor of educational psychology with a longstanding research interest in the psychological and educational problems of the disadvantaged, I was naturally most interested in seeing the Berkeley schools' desegregation program implemented by research into its workings and effects. At a subsequent meeting, the advisory committee decided that I should direct the evaluation study, and in late December 1967, Chancellor Heyns called me to his office to ask me if I would take on the job. I was delighted to do so, although I had misgivings for the short notice, since an essential part of the plan I had suggested was a collection of predesegregation "baseline" data, which meant that the study had to be fully underway in the spring of 1968. The need for this research to get off to a good start by having as thorough an assessment as possible of the schools' educational status before desegregation seemed to me to be of tremendous importance.

Here, for the first time since the momentous Supreme Court decision of 1954; methods for studying desegregation almost as one would conduct an experiment in the laboratory could be developed and applied. I had reviewed the literature on desegregation, and at that time conduct an experiment in the laboratory could be developed and approach what was now possible in Berkeley. White Plains, N.Y., had desegregated its schools the year before, but it was a small community with a much smaller percentage of minority pupils than we had in Berkeley. White Plains' small study of their first year's effects of desegregation had no adequate baseline from predesegregation achievement norms against which to measure changes, and methodologically the study probably could not get by a critical editorial board of any of the top psychological and educational journals. A more thorough effort had gotten underway in Riverside, Calif., the same year, but again, it was a smaller community with far fewer minority pupils, and school integration was not accomplished by busing nor was it as complete as was planned for Berkeley. Also, like White Plains, Riverside had not obtained presegregation baseline data, which seemed to me an indispensable part of the evaluation. In the Berkeley situation, I was convinced, we had all the necessary ingredients for the first major study of the scholastic effects of desegregation. I was determined to make the most of this opportunity by designing the most scientifically impeccable study that time and resources would permit by that late date. The study, if properly done, I believed, could be a model for other communities on the verge of desegregating their schools, and the information gained from the Berkeley experiment could be of immense value to the Nation's schools. The Berkeley schools were pioneer-

ing in school desegregation and there would undoubtedly be valuable lessons to be learned from their experiences. Berkeley is the first city of over 100,000 population to institute complete desegregation and equal proportional representation of all racial and socioeconomic groups in all of its public schools by means of two-way busing. Both majority and minority children are bused from their own neighborhoods to schools which, prior to desegregation, were predominantly either white or black. In 1967 the Berkeley schools enrolled approximately 50 percent minority children; about 40 percent are black.

AN EVALUATION PROPOSAL

The Berkeley Board of Education wanted a contract with the University for doing the evaluation study. Before the contract could be officially drawn up, a definite and explicit research proposal and budget for the first year of operation had to be prepared. This I did in January 1968. My chief aim was to bring whatever expertise I had as a research psychologist and psychometrician to bear on the problems of evaluating the psychological and educational aspects of school desegregation. I had examined previous efforts with a critical eye and was determined to do a much better job in Berkeley than any I had seen elsewhere, even given the modest financial resources made available by the Berkeley School Board.

Certain minimal requirements had to be met if the project was to be of any scientific worth at all, and these I absolutely insisted upon. Adequate baseline data against which future change could be measured did not exist, and to get such data was our first major task. It had to be completed by the end of the spring term of 1968, prior to desegregation the following September. The chief aim of the study was to discover the most significant sources of variance in school performance and pupil behavior. The major sources of variance stem from pupil characteristics (individual differences in measured abilities and attitudes toward school, family background factors and past educational history), classroom climate (peer relationship, class size, the particular admixture and interaction among socioeconomic, racial, and ability levels in the classroom), and the characteristics of the various instructional programs instituted by the schools.

Because of the way classes in the Berkeley schools were constituted, random selection of intact classrooms for testing could not have yielded a representative sample of the school population. Furthermore, in anticipation of the longitudinal study over a period of 5 years or more, it was necessary that we obtain data on as many elementary schoolchildren as possible. The 30 percent mobility rate in the Berkeley schools meant that all available primary pupils should be tested in order to insure a sufficient number of children who could be followed longitudinally through the first 5 years of the new program. Therefore, it was decided that the basic assessment procedures be administered to all pupils in grades K through 6.

OUTSIDE TESTERS

The one thing I insisted on above all was the need for outside testers. The validity of the study, I thought, could be regarded as suspect by

77-942 O—72——34

its critics if the testing were left up to the teachers and other personnel employed by the school district. The essence of an independent evaluation, it seemed to me, was that the basic data be collected by an agency independent of the schools under the most highly standardized conditions feasibly possible. Such conditions cannot be expected to prevail if classroom teachers are called upon to administer tests to their own classes. Consultation with school authorities and with experts in the Educational Testing Service forced me to the conclusion that the only way we could obtain trustworthy data of the kind that a study such as this requires was by having a team of specially trained testers do the entire job of administrating group tests in each classroom. This is essential for the credibility and reproducibility of the data and for having a clear knowledge of the standardized conditions under which they were obtained. Aside from any question of the merits or weaknesses of the tests being used, it is essential that one know precisely the procedures through which performance records were obtained. The reduction in "error variance" and the increase in confidence in the data are fully worth the cost of standardized administration. Another technical reason for controlling the administration of tests is that it permits the investigator to determine the percentage of variance in test scores attributable to differences between testers. We also planned to have approximately half of our testers Negroes and half of them whites, assigned to predominantly Negro and white schools in the counterbalanced fashion of standard experimental design so that the effects of the race of the testers could be assessed for white, Negro, and Oriental children. We were successful in accomplishing this.

It was over this issue of using testers from outside the school system that the project almost succumbed even before it got started. Some of the school authorities were strangely opposed to it, but their reasons convinced me even more of the necessity of using our own crew of testers. I was not concerned, as were some school officials, with how the Berkeley schools' scores might compare with other California school districts whose testing was done by the teachers, nor did I care if our results were out of line with the previous year's test results obtained from teacher-administered tests. I was most concerned that we in fact knew just how the tests had been administered in our study and could therefore do it in the same way in each succeeding year of the evaluation. Since comparisons of test results between different communities which vary in a host of demographic variables related to scholastic achievement are patently ridiculous anyway, I had little sympathy with those whose greatest concern seemed to be how Berkeley's average achievement scores would compare with those of neighboring communities. Unfortunately, my hard-nosed insistence on this point did not make me many friends in the school administration. But at a meeting in the chancellor's office of the advisory committee attended by a number of the school officials, a board member, and myself, it became absolutely clear that my continuing as director of the project was contingent upon doing the testing as I thought best, and the school officials finally agreed to it. For the purposes of a study such as this, test data obtained in the usual way would not be of much interest to a researcher.

We trained 15 persons, young men and women, half of them white and half of them Negro, to administer the battery of tests to the 9,000

elementary schoolchildren in Berkeley during April and May 1968. All the testers were college graduates, some with M.A.s and Ph.D.s, several with a teaching background. They were recruited through the University's employment bureau and through various civic organizations in the community, such as the League of Women Voters and the Urban League. Throughout the course of the testing in the schools, a supervisor visited classes to observe each tester in action to help issue a kind of "quality control" of the testing procedures. I personally met with the teachers and staff in the elementary schools to explain the purpose of the study and answer their questions.

The results of the achievement testing especially indicated that the effort to minimize variations in testing procedures by having all tests administered by a staff trained to follow the same procedures was highly worthwhile. Judging from the test results of previous years, it appears that the effects of procedural variations in testing by the teachers tend to cancel out, leaving the overall mean at each grade level and in each school very much the same as when we administered the tests ourselves. But the error component in individual scores is considerably greater when testing procedures are not controlled. This was reflected in the large and significant decrease in standard deviations of achievement test scores from 1967 and 1968. The decreases in SD are significant well beyond the .01 level and furthermore represent a very substantial decrease in the total test variance, which is presumably a reduction in error variance. For example, in total reading scores (of the Stanford achievement battery), the variance decreased from 1967 to 1968 by 23 percent for grade 1, 28 percent for grade 2, and 23 percent for grade 3. This degree of reduction in error variance can be a considerable advantage in increasing the precision of the statistical and correlational analysis of these data.

INDIVIDUAL TESTING

In order to evaluate the effects of classroom atmosphere on test performance and to estimate the validity of group testing in general, and particularly for minority children, it was decided to administer individually each of the ability and achievement tests in our battery to a random sample of the school population. The children selected for individual testing were not given the same test in a group procedure, and no child was given more than one of the tests individually. A random sample of all K–6 pupils in the 14 schools was accomplished by taking one child at random from each class at the same time the group test was administered and giving that child individually the same test that was being administered to the whole class. The randomly selected children were tested individually in a private room in the school or in a testing van. It was found that, on the average, children do slightly less well when tested individually than when tested in a group in the classroom, although the average difference amounts to only about 10 to 15 percent of a standard deviation (equivalent to 2 IQ points)—slightly more for Orientals and Negroes and slightly less for whites.

GENERAL POLICIES

A general policy was laid down that it should be made clear to all concerned that the data of the project were to be collected only for

statistical and research purposes. Any personal information obtained from pupils, teachers, or parents was to be the confidential possession of the project, and teachers and parents were assured that any data obtained in interviews, questionnaires, and classroom observations would strictly preserve the respondents' anonymity. I also insisted that there be no restrictions on the types of data collected or their method of collection, so long as these practices are ethical, are completely open to public inspection, are scientifically justified, and are in no way physically or psychologically harmful to the participants. No tests or other evaluative procedures were conducted without fully informing the superintendent of schools of the nature and purpose of the proposed procedures. It was agreed, however, that information of a personal nature, obtained in interviews, in opinion questionnaires, and in attitude, personality, and adjustment inventories, would never be obtained by the project without the written consent of the superintendent, who reserved the right to veto without question the use of any particular instrument of this type.

Also, in a project of this type, the school authorities should formally consent to a maximum time limit for pupil participation in the evaluation procedures, and all such procedures should be carefully budgeted within this time limit. Researchers need to be assured that they can plan a proper study in terms of available test time and access to classrooms. But also there must be reasonable restrictions on the amount of pupils' in-school time that can be used for purposes of research and evaluation. I proposed that the schools agree to the following time-budget: (*a*) The mean testing time per child throughout all grades would never exceed 0.75 percent of the child's total in-school time per school year, and (*b*) the maximum amount of time for any single child's participation should never be allowed to exceed 1 percent of his in-school time per year.

KINDS OF DATA OBTAINED

A variety of data were desired for the baseline study, both as a means of describing the Berkeley elementary school population just prior to desegregation and as a basis for future comparisons. The types of assessment were as follows:

A. Ability testing (grades K–6).

B. Motivational assessment (grades 1–6).

C. Scholastic achievement testing (grades 1–6).

D. Sociometric assessment of peer relationships with classrooms (grades 4–6).

E. Teacher attitude inventory (all teachers of grades 4–6).

F. Parent opinion assessment (questionnaires to parent or guardian of all children in grades K–6 on a voluntary basis).[1]

G. Assessments of pupil behavior and classroom climate by means of standardized observational techniques.

H. Supplementary background data.

The instruments and methods used for obtaining these data are described in the appendix.

Probably the most innovative aspect of the research design was to make use of sibling data. The most powerful method of statistically

[1] The results of this questionnaire have been published (see No. 2 below).

controlling for differences in family background variables is by comparing the scholastic achievements of younger years in segregated schools; younger siblings will come up though the grades in integrated schools. Thus comparisons of performance under segregated and integrated classes can be made with considerable precision when we can statistically control most of the variance associated with family background factors. This can be done most efficiently, not by obtaining a great amount of detailed personal information about the child's parents or home conditions, but simply by being able to identify all of the child's siblings who are also in school and whose test scores are also a part of the data bank. The more precise and complete this information is, the sharper can be the analysis of the data, so that even quite small changes in scholastic achievement can be detected as statistically significant, since they are not swamped by the major sources of variance over which the schools themselves have little or no direct control. A number of important questions were posed which the study was designed to answer concerning the progressive and cumulative effects of desegregation on scholastic performance, classroom behavior and morale attitudes of pupils, parents, and teachers and demographic changes in the community during the first 5 years of desegregation.

All of the proposed baseline data were collected during spring 1968, before any part of the elementary school integration plan had been put into effect. Only our staff of 20 dedicated workers will ever really appreciate the effort, patience, ingenuity, and diplomacy it required to obtain this large amount of data on some 9,000 children and their teachers and parents in the space of 2 months. It could not have been possible without the organizational and diplomatic talents of my two tireless chiefs of staff, Dr. Wade Egbert and Mrs. Alice Carter. The scheduling of all this testing in 14 schools under conditions that fluidly change from day-to-day, requiring frequent last-minute rescheduling, was at least as complicated as playing simultaneous chess on 20 boards, and often much more frustrating. Considering the time pressure we were under, and the tensions apparent in many of the Berkeley schools immediately preceding the drastic changes that were soon to come about, I still marvel at how my staff managed to accomplish all this massive data collection so successfully. These data undoubtedly comprise the most complete set of baseline measurements ever undertaken for a study of desegregation. Indeed, they are among the most thorough assessments ever made of a large elementary school population for any purpose. The analyses of most immediate interest to the school administration were carried out by computer techniques and summary reports were turned over to the school authorities. All the basic data have remained in possession of the university, have been coded on computer tapes, and are still undergoing statistical analysis directed at answering a variety of research questions. Most of these studies will eventually be published in the appropriate psychological and educational journals.

It is of interest to know something of the climate of public opinion in Berkeley just prior to the enactment of desegregation. Berkeley is, of course, not a typical American city, and, as one would imagine, only an unusually liberal and progressive community could be expected to take the lead in such an endeavor as the complete racial and socioeconomic desegregation of its public schools. Our parent and

teacher opinion questionnaire gives a fair picture of specific attitudes toward desegregation in spring, 1968. Completed questionnaires concerning opinions regarding racial integration, busing, and ability grouping were obtained from 337 Berkeley elementary school teachers (with 71 percent returns) and from the parents of over 8,000 elementary school pupils (with 4,596 parents responding).[8] Analysis of the questionnaire results indicate that: (1) the vast majority of Berkeley teachers favored integration and busing and held attitudes favorable to the school administration's official policies in this area; (2) older teachers were less favorably disposed toward integration and busing than younger teachers; (3) the majority of parents favored integration but were less unanimous in their approval of busing as a means of achieving integrated schools; (4) more females than males favored busing; (5) there were significant racial differences in opinions on busing, with Negroes most favorable, Orientals least, and whites intermediate; (6) a majority of all racial groups favored ability grouping; (7) favorable attitudes toward busing decreased with number of years residence in Berkeley; (8) homeowners approved of busing less than renters; and (9) favorableness toward integration and busing was positively related to parents' educational level.

OBSTACLES AND DIFFICULTIES

Since we have no adequate basis for comparison it is practically impossible to know if the various problems we encountered were more or less than would have occurred in any other school system faced with the same invasion of testers, demands of scheduling, and so forth. The time pressure no doubt made it trying for everyone concerned. But there were also problems that were of a political-ideological nature. For a community of its size, Berkeley has an unusual number of highly vociferous politically active groups which constantly badger the school board and the administration over almost every step they take. This reality, I felt, was often uppermost in the minds of many of the school officials I had to deal with in planning and executing the study. As a consequence, the administrations motives, often governed by public relations considerations, and mine, governed by considerations of proper research methodology, were at time in conflict. These conflicts were almost always resolved by compromise or persuasion. But I never compromised on any point that I thought could impair the scientific integrity of the study. I was aware that my stubbornness in this regard rubbed some persons the wrong way, but I was more concerned with doing a good study than with being well-liked by everyone concerned. My need to be liked by people in general is perhaps peculiarly low in my hierarchy of personal values. Those who tried to read political-ideological motives into my research proposal, and on these grounds may have felt some suspicion or mistrust of my role in their midst, simply do not know me or my motivations. I am even somewhat ashamed to admit being such a nonpolitical creature as in fact I am. I have little patience with those who view scientific research as a mere

[8] A complete report of these data has been published: Jensen, A. R. Parent and Teacher Attitudes Toward Integration and Busing, *Research Résumé*, No. 43, California Advisory Council on Educational Research, 1970.

vehicle for propagandizing some particular ideology, and I imagine there may be times when my disdain for this philosophy is not well concealed from the ideologues. I heard it asked, in all seriousness, by one person at a school meeting, for example: "Since we know that research results can be made to show anything we want, why not make them show what we believe in, and what we're trying to work for?" This deplorable philosophy, I fear, has become more common among some highly politicized educators and social scientists in recent years.

There were also the difficulties inherent in dealing with liaison persons who themselves were not in authority and therefore could not deal directly with various problems that arose. For this reason I usually found it easiest of all to deal with Dr. Neil Sullivan, the superintendent of schools. There was never any doubt of his authority. He clearly ran the show. And he had to be admired for the way he took on a tremendously difficult goal in Berkeley and brought it to reality with éclat and aplomb. He seemed to me intelligent and competent, tough and shrewd—traits which I find make a person easy to respect and to work with even when there are disagreements on specific issues. It was clear to me that the picture many of the public had of Sullivan as a mere sentimentalist was all wrong. Such a person, of course, could not have done the difficult job that Sullivan accomplished. I was never in disagreement with Sullivan over major objectives, but we differed at times in certain specific aspects of strategy and tactics. Considering some of the thorny problems posed by the whole idea of making an independent evaluation of Sullivan's program, it all went along quite well. Sullivan understandably had his particular concerns as a superintendent and I had mine as a researcher. It required some astuteness from both sides to meet the legitimate needs of each. I never doubted Sullivan's astuteness in this respect.

POLITICALLY MOTIVATED CRITICISMS

Quite late in the course of our data collection, a political activist group, the Community for New Politics (CNP) sent a spokesman to the public board meeting of May 7, 1968 to protest against the evaluation study. A list of criticisms of the study was presented which questioned the relevance of some of the tests, questionnaires, and other information we were obtaining. The accusations and criticisms were so surprisingly false and technically incompetent that one can suspect that their authors did not expect that any rebuttal would ever be made. A large part of the CNP statement was directed against me personally, alleging racist biases and questioning my impartiality in evaluating the educational effects of school desegregation. All these points I countered head-on in my rebuttal at the very next public meeting of the board on May 21, 1968.[3]

DEMISE OF THE FOLLOWUP STUDY

My outspoken views, expressed at professional and scientific meetings and in journal articles, to the effect that all of the causes of racial

[3] Copies of the CNP statement and Jensen's rebuttal are available from the author.

differences in mental abilities and scholastic achievements were not yet scientifically established and that the probable role of genetic factors had not been adequately researched, were publicized quite luridly (once there was even a 2-inch banner headline printed in red) in the local newspapers soon after I became associated with the Berkeley school evaluation project. This aroused a good deal of opposition toward me in some circles, which, from the standpoint of political and public relations appearances, undoubtedly made me less than ideal as a director of the project. Many of the school administrative personnel thought I was too controversial to be in this position. I could hardly disagree, considering some of the misleading publicity I had received. Although neither I nor any of my colleagues ever doubted my scientific integrity in the project, it was recognized that the same qualities that I considered a virtue as a researcher, and which made me outspoken in my writings about important educational issues, also were the very qualities that made me so unacceptable to some politically oriented critics. This problem was fully and openly discussed with the school authorities and with the university advisory committee. It was decided, with my full approval (although I probably really had no choice in the matter), that beyond the collection and analysis of the baseline data, the project would be nominally headed by the dean of the school of education as the principal investigator. There were to be two other administrative positions under him—director and program coordinator—assigned to persons not previously identified with this project or even with this field of educational research, thereby preserving a kind of anonymity at the top of the project which is most visible to the media and the public. It was intended that I should recede far into the background, as research psychologist on the project, hidden from the public firing line but remaining in charge of all the psychometric testing and analysis. I, along with several researchers on the U.C. campus in the department of education, sociology, and psychology, worked together during the summer of 1968 to prepare an elaborate research proposal intended to describe the project in detail through the first 5 years of the desegregation program. Our hope was to use the proposal to obtain the necessary funding from one of the large Federal or private agencies that make grants for educational research. While these possibilities were being sought and explored during the ensuing months, the project, which up till then had been funded by the Berkeley school district and the university, came to its sudden demise, just about 2 weeks after we had begun the spring, 1969 testing intended to assess the first year of the new program.

The precise chain of events that led to this (to me) sudden announcement remains rather obscure, at least to me. In the first week of March 1969 the *Harvard Educational Review* (winter 1969) published my lengthy article, "How Much Can We Boost IQ and Scholastic Achievement?" It was quickly publicized in the popular press, at times with highly misleading distortions, and was especially aired by the newspapers and radio in the Bay area. Some of the political pressure groups in Berkeley were incensed by these popular accounts of my views and clamored for my being fired, not only from my role in the evaluation project, but even from my position as a professor in the university. A spokesman for one group proposed at an open meeting of

the school board that the evaluation study be immediately halted and that all the baseline data we had collected the previous year be destroyed. When the president of the board pointed out that all the data were in possession of the university and that therefore the board had no jurisdiction over them, the board was then urged to dissociate completely from the project. (A few weeks before this eruption, I had been told by Berkeley's present school superintendent, Dr. Richard Foster, that when he first took office, early in 1969, there were people in the community who urged that the first thing he should do as the new superintendent is "get rid of Jensen.") Unfortunately, at the very next public meeting of the board, it announced its decision that the Berkeley schools would dissociate itself from the university's evaluation project. Our testers, who were already in the schools, were retained to complete the spring testing, but all the test results were kept by the school administration for scoring and analysis and the project at the university closed up shop. I was notified of the board's decision the next day by the head of the project, the dean of the school of education. The same day a school official explained to me that they were reluctantly forced to come to this decision because the schools were not a research institute, but a political unit, and they had to be sensitive to the political climate of the community. This then, was given as at least part of the reason for the demise of the project. The school administration has carried on with the routine State-mandated testing, but, as far as I know, nothing resembling the thorough and comprehensive research we had planned has been attempted, and I have not found any of the routine reports presented at board meetings very interesting or informative from a research standpoint. Thus, a real assessment of the effects of total school desegregation still remains an unaccomplished task for educational research. Perhaps it can be done, somewhere, at some time in the future. It could still be done in Berkeley, because all the predesegregation baseline data are completely intact.

A PERSONAL VIEW OF THE ISSUES

From the beginning, I personally favored racial desegregation of schools, and I hopefully believe it is coming about throughout the Nation. As an educator, I am concerned that it come about in such a way as to be of benefit to the schooling of all children. Achieving racial balance, while viewed by many of us as desirable for moral, ethical, and social reasons, will not solve existing educational problems; it will create new ones, and I am anxious that we provide the means for fully and objectively assessing them and for discovering the means of solving them. I am quite convinced on the basis of massive research evidence that the educational abilities and needs of the majority of white and Negro children are sufficiently different at this present time in our history that both groups—and particularly the more disadvantaged group—can be cheated out of the best education we now know how to provide in our schools if uniformity rather than diversity of instructional approaches becomes the rule. Diversity and desegregation need not be incompatible goals. I think both are necessary. But achieving racial balance and at the same time ignoring individual differences in children's special educational needs could be most destructive to those who are already the most disadvantaged educationally. The allocation

of a school's resources for children with special educational problems cannot be influenced by race; it must be governed by individual needs. Making an association, as some persons do, between the "nature-nurture" (or "heredity-environment") question and the issue of racial desegregation of schools is, to me, a most flagrant *non sequitur*. The pros or cons of school integration have no logical or necessary connection with the question whether there are or are not racial genetic differences in mental ability, and the outcome of scientific research on this legitimate question should have no bearing, either one way or the other, on the issue of school integration.

Since educators have at least officially assumed that race and social class differences in scholastic performance are not associated with any genetic differences in growth rates or patterns of mental abilities but are due entirely to discrimination, prejudice, inequality of educational opportunity, and factors in the child's home environment and peer culture, we have collectively given little if any serious thought to whether we would do anything differently if we knew in fact that all educational differences were not due solely to these environmental factors.

There have been and still are obvious environmental inequities and injustices which have disfavored certain minorities, particularly Negroes, Mexican-Americans, and American Indians. Progress has been made and is continuing to be made to improve these conditions. But there is no doubt there is still a long way to go, and the drive toward further progress in this direction should be given top priority in our national effort.

Education is viewed as one of the chief instruments for approaching this goal. Every child should receive the best education that our current knowledge and technology can provide. This should not imply that we advocate the same methods or the same expectations for all children. There are large individual differences in rates of mental development, in patterns of ability, in drives and interests. These differences exist even among children of the same family. The good parent does his best to make the most of each child's strong points and to help him on his weak points but not make these the crux of success or failure. The school must regard each child, and the differences among children, in much the same way as a good parent should do.

I believe we need to find out the extent to which individual differences, social class differences, and race difference in rates of cognitive development and differential patterns of relative strength and weakness in various types of ability are attributable to genetically conditioned biological growth factors. The answer to this question might imply differences in our approach to improving the education of all children, particularly those we call the disadvantaged, for many of whom school is now a frustrating and unrewarding experience.

Individuals should be treated in terms of their individual characteristics and not in terms of their group membership. This is the way of a democratic society, and educationally it is the only procedure that makes any sense. Individual variations within any large socially defined group are always much greater than the average differences between groups. There is overlap between groups in the distributions of all psychological characteristics that we know anything about. But dealing with children as individuals is not the greatest problem. It is

in our concern about the fact that when we do so, we have a differentiated educational program, and children of different socially identifiable groups may not be proportionately represented in different programs. This is the "hangup" of many persons today and this is where our conceptions of equal opportunity are most likely to go awry and become misconceptions.

Group racial and social class differences are first of all individual differences, but the causes of the group differences may not be the same as of the individual differences. This is what we must find out, because the prescription of remedies for our educational ills could depend on the answer.

Let me give one quite hypothetical example. We know that, among middle-class white children, learning to read by ordinary classroom instruction is related to certain psychological developmental characteristics. Educators call it "readiness." These characteristics of readiness appear at different ages for different kinds of learning, and at any given age there are considerable individual differences among children, even among siblings reared within the same family. These developmental differences, in middle-class white children, are largely conditioned by genetic factors. If we try to begin a child too early in reading instruction, he will experience much greater difficulty than if we waited until we saw more signs of "readiness." Lacking readiness, he may even become so frustrated as to "turn off" on reading, so that he will then have an emotional block toward reading later on when he should have the optimal readiness. The readiness can then not be fully tapped. The child would have been better off had we postponed reading instruction for 6 months or a year and occupied him during this time with other interesting activities for which he was ready. Chances are he would be a better reader at, say, 10 or 11 years of age for having started a year later, when he could catch on to reading with relative ease and avoid the unnecessary frustration. It is very doubtful in this case that some added "enrichment" to his preschool environment would have made him learn to read much more easily a year earlier. If this is largely a matter of biological maturation, then the time at which a child is taught in terms of his own schedule of development becomes important. If, on the other hand, it is largely a matter of preschool environmental enrichment, then the thing to do is to go to work on the preschool environment so as to make all children equally ready for reading in the first grade. If a child's difficulty is the result of both factors, then a combination of both enrichment and optimal developmental sequencing should be recommended.

There is a danger that some educators' fear of being accused of racial discrimination could become so misguided as to work to the disadvantage of many minority children. Should we deny differential educational treatments to children when such treatment will maximize the benefits they receive from schooling, just because differential treatment might result in disproportionate representation of different racial groups in various programs? I have seen instances where Negro children were denied special educational facilities commonly given to white children with learning difficulties simply because school authorities were reluctant to single out any Negro child, despite his obvious individual needs, to be treated any differently from the majority of youngsters in the school. There was no hesitation about singled out

white children who needed special attention. Many Negro children of normal and superior scholastic potential are consigned to classes in which one-fourth to one-third of their classmates have IQs below 75, which is the usual borderline of educational mental retardation. The majority of these educationally retarded children benefit little or not at all from instruction in the normal classroom, but require special attention in smaller classes that permit a high degree of individualized and small group instruction. Their presence in regular classes creates unusual difficulties for the conscientious teacher and detracts from the optimal educational environment for children of normal ability. Yet there is reluctance to provide special classes for these educationally retarded children if they are Negro or Mexican-American. The classrooms of predominantly minority schools often have 20 to 30 percent of such children, which handicaps the teacher's efforts on behalf of her other pupils in the normal range of IQ. The more able minority children are thereby disadvantaged in the classroom in ways that are rarely imposed on white children for whom there are more diverse facilities. Differences in rates of mental development and in potentials for various types of learning will not disappear by being ignored. It is up to biologists and psychologists to discover their causes, and it is up to educators to create a diversity of instructional arrangements best suited to the full range of educational differences that we find in our population. Many environmentally caused differences can be minimized or eliminated, given the resources and the will of society. The differences that remain are a challenge for public education. The challenge will be met by making available more ways and means for children to benefit from schooling. This, I am convinced, can come about only through a greater recognition and understanding of the nature of human differences.

But research in this realm, unlike most scientific research, is fraught with difficulties not intrinsic to the research itself. Because I have refused to assume a doctrinaire stance regarding the roles of genetic and environmental factors in the causation of racial differences in intelligence and scholistic performance, I have become persona non grata to the dogmatists at both extremes of this issue. But the personal and professional consequences of my dragging these important questions out from under the rug and suggesting that they be subjected to scientific study are another story.

In a free society, one which permits freedom of speech and of the press, both to express and to criticize diverse views, the social responsibility of the scientist, it seems to me, is perfectly clear. It is simply to do his research as competently and carefully as he can, and to report his methods, results, and conclusions as fully and as accurately as possible. The scientist, when speaking as a scientist about his research, should not make it subordinate to his nonscientifically arrived-at personal, social, religious, or political ideologies. We have seen clear examples of what happens when science is corrupted by servitude to political dogma—in the bizarre racist theories of the Nazis and in the disastrous Lysenkoism of the Soviet Union under Stalin. Unfortunately, we have been witnessing in the United States similarly ideologically motivated dogmatism concerning the causes of obvious differences in average educational and occupational performance

among various subpopulations, socially identified as racial groups. Serious consideration of the question of whether the observed racial differences in mental abilities and scholastic performance involve genetic as well as environmental factors has been totally taboo in academic, scientific, and intellectual circles in the United States in recent years. Nevertheless, it remains a persistent question. My belief is that scientists in the appropriate disciplines must finally face the question squarely and not repeatedly sweep it back under the rug. In the long run, the safest and sanest thing we can urge is intensive, no-holds-barred inquiry in the best tradition of science.

There is perhaps an understandable reluctance to come to grips scientifically with the problem of race differences in intelligence—to come to grips with it, that is to say, in the same way that scientists would approach the investigation of any other phenomenon. This reluctance is manifested in a variety of "symptoms" found in most writings and discussions of the psychology of race differences, particularly differences in mental ability. These symptoms include a tendency to remain on the remotest fringes of the subject; to sidestep central questions; to blur the issues and tolerate a degree of vagueness in definitions, concepts, and inferences that would be unseemly in any other realm of scientific discourse. Many writers express an unwarranted degree of skepticism about reasonably well-established quantitative methods and measurements. They deny or belittle already generally accepted facts—accepted, that is, when brought to bear on inferences outside the realm of race differences—and they demand practically impossible criteria of certainty before even seriously proposing or investigating genetic hypotheses, as contrasted with extremely uncritical attitudes toward purely environmental hypotheses. There is often a failure to distinguish clearly between scientifically answerable aspects of the question and the moral, political, and social policy issues; a tendency to beat dead horses and to set up straw men on what is represented, or misrepresented, I should say, as the genetic side of the argument. We see appeals to the notion that the topic is either too unimportant to be worthy of scientific curiosity or too complex, or too difficult, or that answers to key questions are fundamentally "unknowable" in any scientifically acceptable sense. Finally, we often see the complete denial of intelligence and race as realities, or as quantifiable attributes, or as variables capable of being related to one another—in short, there is an ostrich-like dismissal of the subject altogether.

I believe these obstructive tendencies will be increasingly overcome the more widely and openly the subject is researched and discussed among scientists and scholars, and the more the general public is made intelligently aware of the issues. As some of the taboos against open discussion of the topic fall away, the issues will become clarified on a rational basis. We will come to know better just what we do and do not yet know about the subject, and we will be in a better position to deal with it objectively and constructively.

We must always distinguish clearly between research on racial differences and racism. Racism usually implies hate or aversion and is aimed at the denial of equal rights and opportunities to persons on the basis of their racial origin. Racism should be attacked directly in

the spheres in which it operates, by enacting and enforcing laws and arrangements that help to insure equality of civil and political rights and to guard against discrimination in educational and occupational opportunities on the basis of racial membership.

To fear research on genetic racial differences, or the possible existence of a biological basis for differences in abilities, is, in a sense, to grant the racist's assumption—that if it should be established beyond reasonable doubt that there are genetically conditioned differences in mental abilities among individuals or groups, then we are justified in oppressing or exploiting those who are most limited in genetic endowment. This, of course, does not follow at all. Equality of human rights does not depend upon the proposition that there are no genetically conditioned individual differences or group differences. Equality of rights is a moral axiom: it does not depend upon any set of scientific data.

I have always advocated dealing with persons as individuals, each in terms of his own needs and characteristics, and I am opposed to according to persons on the basis of their race, color, national origin, or social class background. But I am also opposed to ignoring or refusing to investigate the causes of the well-established differences among racial groups in the distribution of educationally relevant traits, particularly IQ.

Present educational programs generally are failing to provide all segments of our population with the knowledge and skills needed for economic self-sufficiency in our increasingly technological society. Literal equality of educational opportunity falls short of solving this problem. In appropriate instructional procedures, often based on the notion that all children can learn best in essentially the same way except for easily changed environmental influences, can alienate many children from ever entering upon any path of educational fulfillment.

In our efforts to improve education we should not lose sight of the focal point of our concern—the individual child. And this means the biological as well as the social individual, for man's intelligence and educability are the products of biological evolution as well as of individual experience. I believe that, in the long run, the greatest respect educators can pay the children in our schools is to try to take full account of all the facts of their nature. In my view, more thought and effort should be directed into discovering ways of modifying our educational methods and goals to accord with the full variety and range of children's abilities and proclivities. In that way, perhaps, many more children can truly profit from their years in school.

APPENDIX

A. ABILITY TESTING

1. Lorge-Thorndike Intelligence Tests (Grades K–6)

From the standpoint of test construction and standardization this is probably the best group intelligence test on the market. An added advantage is that it has been used in past years in the Berkeley schools and is the most generally used test in other school systems in the bay area. Also, more is known about socioeconomic and educational correlates of the Lorge-Thorndike tests than is known for any other group-administered tests.

2. Tests of attention and memory span (Grades 2–6)

These tests, which together take less than 30 minutes to administer, had been undergoing research and development in the schools of Richmond and Orinda in 1966–67. Their chief value lies in the fact that they are quite "culture fair," they assess abilities important for classroom learning, and they reveal able learners in culturally disadvantaged groups who often go undetected by means of the usual standardized intelligence tests. These tests reliably measure an important aspect of learning ability which is not measured by traditional tests. The attention and memory span tests are administered by means of a tape recording.

3. Figure copying test (Grades K–4)

This test was developed at the Gesell Institute at Yale University. It is intended as a means for measuring developmental readiness for the traditional tasks of the primary grades. The child is simply asked to copy 10 geometric forms. There is no time limit, but the test usually takes less than 15 minutes to administer in the classroom.

B. MOTIVATIONAL ASSESSMENT (GRADES 1–3)

1. Test of speed and persistence

This is a simple test that takes at most 10 minutes to administer. The child simply makes *X's* in a series of squares for the two periods, each of 1–5 minutes' duration, separated by a 2-minute rest period. The test is valuable because it contains little or no intellectual component. The pupil's score reveals his ability to mobilize his effort for a brief period and to persist in an assigned task. It serves partly as an index of classroom morale and can be used as a means of estimating the child's test-taking attitude and effort: It can be entered as a moderator variable into correlational analyses among other ability and achievement tests.

C. SCHOLASTIC ACHIEVEMENT TESTING (GRADES 1–6)

1. Stanford achievement tests

These are the best standardized tests of scholastic achievement and have the added advantage that they have been in general use in the

Berkeley schools in past years. Because of the limited time, only the "partial battery" was used, consisting of word meaning, paragraph meaning, spelling, language, word study skills, arithmetic, computation, arithmetic concepts, and arithmetic applications.

D. SOCIOMETRIC ASSESSMENT (GRADES 4–6)

This brief questionnaire can be administered by the teachers. It is a typical sociometric device: the child answers a number of simple questions about his relationship to his classmates. The form proposed for use here has been developed by Dr. Nadine Lambert of the Department of Education of the university. It had also been used in the Riverside Schools evaluation. It is probably the most objective indicator of the organization of peer relationships in the classroom. Postintegration assessments should reflect the extent to which integration is actually taking place at the classroom level in the relationships among children. It indicates the degree to which interpersonal relationships follow ethnic, social-class, and ability factors as they are perceived in the classroom. It was given only in Grades 4–6 because it requires a level of reading and writing ability not generally found much below this grade level.

E. TEACHER ATTITUDE

A brief inventory of teacher attitudes and opinions regarding integration, busing, ability grouping, etc. was adapted from already existing inventories. It was distributed to teachers who mailed their completed questionnaires directly to the project director. Only a statistical summary of the responses was made available to school officials. Individual teachers remained strictly anonymous.

F. PARENTS ATTITUDES AND OPINIONS

This brief questionnaire was adapted from the questionnaires developed by Dr. Leonard Marascuilo of the department of education at the university for use in previous studies of public opinion regarding school integration, busing, et cetera. Questionnaires were sent to parents and their replies were mailed by them directly to the project director. Reply was voluntary and anonymous.

G. PUPIL BEHAVIOR AND CLASSROOM CLIMATE

These factors were assessed by observational techniques developed by Dr. Theodore Parsons of the University of California. Five trained observors, using already developed checklists and rating forms made observations in all elementary classrooms in the system.

H. SUPPLEMENTARY BACKGROUND DATA

Certain items of information not routinely entered in the child's school record were obtained for a more adequate picture of demographic characteristics of the Berkeley school population.

(1) Birthdate and present age.

(2) Sex.

(3) Race.
(4) Race of father, race of mother.
(5) Father's occupation, mother's occupation.
(6) Highest school grade completed by father.
(7) Highest school grade completed by mother.
(8) Mother's birthdate (mother's age is likely to be a significant source of variance in school performance).
(9) Home address.
(10) Position in family (birth order).
(11) Number of half-siblings and their ages.
(12) Number of full-siblings and their ages.
(13) Number of step-siblings and their ages.
(14) Complete identification of all full- and half-siblings by name.
(15) Number of adults (over 18) in household.
(16) Child's height and weight.
(17) Whether left or right handed; left or right eyed.
(18) For grades K and 1 only: missing baby teeth.
(19) Whether the child has been in special programs, such as Head Start, busing, et cetera.
(20) Language spoken in the home.

77-942 O—72——35

ON THE EXPLANATION OF RACIAL-ETHNIC GROUP DIFFERENCES IN ACHIEVEMENT TEST SCORES

By George W. Mayeske,* U.S. Office of Education, Washington, D.C.

INTRODUCTION

My colleagues[1] and I in the U.S. Office of Education have been privileged to be charged with the responsibility for illustrating and documenting the structure and functioning of the American public school system. We could not dream of having such a lofty objective if we did not have at our disposal the most comprehensive body of data ever collected on public schools and their students in the United States. I am, of course, referring to the educational opportunities survey data collected in the fall of 1965 at the direction of Congress in the Civil Rights Act of the prior year. A report utilizing this data to investigate the equality of educational opportunity for various racial and ethnic groups was issued in the fall of 1966 under the principal authorship of James S. Coleman. Today I would like to present excerpts from two reports that utilized this same body of data—Mayeske, et al., 1969; 1971—as well as some special analyses that pertain to the explanation of racial-ethnic group differences in achievement. But let me first focus on the nature and scope of the data base and the background work that was done in preparation for these reports.

THE DATA BASE AND BACKGROUND WORK

The educational opportunities survey entailed the testing and surveying of about 650,000 students in some 4,000 public schools, throughout the country in grades 1, 3, 6, 9 and 12, together with their teachers, principals, and superintendents. The survey sample consisted of a 5-percent sample of schools. The data base is comprehensive in that detailed factual and attitudinal information was collected on the students home background, attitude toward school, race relations, and the world. A battery of ability and achievement tests was administered at each grade level. Information was collected from some 60,000 teachers and 4,000 principals concerning their training and experience, their view of the school, et cetera. The final part of the teacher questionnaire consisted of a 30-item contextual vocabulary test which was intended to be a measure of the verbal facility of the teacher. In addition, the principal provided data on the school's facilities, staff, programs, curriculum, et cetera. For further detailed information on the survey data I will refer you to the report "Equality of Educational Opportunity" (Coleman et al., 1966).

---

*The views expressed in this paper are those of the author and do not reflect the official policy of the U.S. Office of Education.

[1] Tetsuo Okada, Carl E. Wisler, Wallace M. Cohen, and Albert E. Beaton, Jr.

The main goal of our background work was to reduce the more than 400 variables in an empirically meaningful way into indexes and sets of indexes so that the volume of data processing and complexity of later analyses could be reduced. Before the variables could be reduced into meaningful groupings, however, decisions had to be made concerning the estimation of missing data and the coding or scaling of variables. As a guide in the estimation of missing data or handling of nonresponses, it was decided to analyze the responses to each question against one or more criteria or dependent variables so that not only the percent responding to each item or response alternative, but also their mean score on the dependent variable could be used as a guide in coding the variables and in assigning a value to the nonrespondents. Since the approach differed somewhat for the student, teacher, and principal questionnaires each analysis will be described separately.

A factor analysis of the achievement measures [2] showed that a single factor could be used to describe their intercorrelations.[3] Accordingly, the weights from the first principal component of the intercorrelations were used to weight scores on the individual tests and sum them to obtain the overall composite of academic achievement.[4] It was this composite which was used as a criterion against which item responses were analyzed (Mayeske, et al., 1968). This composite is also the dependent variable for many later analyses.

In order to maximize the linear relationship of each student variable with student achievement, criterion scaling (Beaton, 1969) was employed. By criterion scaling is meant that each item response was coded or scaled by assigning the mean value of the dependent variable for each of the different response alternatives for an item.[5]

For the teacher variables, each item was analyzed against the teacher's total score on a self-administered contextual vocabulary test (Mayeske, et al., 1967). For the principal variables, each item was analyzed against the number of students enrolled in the school, the rural-urban and socioeconomic status of the school and, the principal's salary (Mayeske, et al., 1968). These analyses were used as guides in assigning codes or scale values and in estimating missing data.[6]

To obtain meaningful groupings of variables, the intercorrelations of the student, teacher and principal sets of variables were each sub-

---

[2] The tests were: (1) General information; (2) reading comprehension; (3) mathematics achievement; (4) verbal ability and; (5) nonverbal ability. For grades 9 and 12 all five tests were available, for grades 6 and 3 tests (2) through (5) were available and, for grade 1 only (4) and (5) were available. See Coleman (1966) for details on these tests.

[3] The first principal component of the intercorrelations accounted for 75 percent of the variance for grades 6, 9, and 12; 60 percent for grade 3; and 82 percent for grade 1. Similar results were obtained for separate racial-ethnic and regional groupings of students (Mayeske and Weinfeld, 1967).

[4] For grades 6, 9 and 12 tests (3) and (5) received slightly lower weights than for the other grade levels. See the appendix for the group standard deviations on the composite for the different grade levels.

[5] Almost all of the student variables were coded in this manner. When the results of this scaling technique were compared with a more conventional procedure it was found that they were very similar except for some of the attitudinal items which were linearized by the criterion scaling procedure (Mayeske, et al., 1969).

[6] However, for the teachers and principals questionnaires the items were not coded so as to maximize their relationship with these dependent or criterion variables.

jected to a series of factor analyses. The principal component technique was used to extract components and the Varimax technique was used to rotate components having a root of one or greater (Horst, 1965). This approach was essentially iterative in that variables that did not form meaningful groupings or blurred and otherwise meaningful grouping were eliminated and the remaining variables were refactored. The teacher and student variables readily fell into meaningful groupings after two iterations which resulted in the elimination of about six to 12 variables from each set. The highest weights from the Varimax rotation were used to combine the variables to obtain index scores. In order to keep the index score intercorrelations low a variable was allowed to have a weight on only one index.

The variables from the principal questionnaire dealt with a wide variety of different aspects of the school. These variables did not readily fall into any naturally meaningful groups. Consequently, a priori groupings, such as variables concerned with the physical plant or instructional facilities were subjected to a principal component analysis. The weights from the first principal component were then used to obtain index scores for each school.

A brief description of the indexes obtained and other variables retained for future analyses are given in the appendix. Time does not permit for a full discussion of them now, however, they are given for reference purposes and will be discussed in later portions of the text. Using these indexes we have conducted an extensive number of analyses, a small portion of which I would like to present today. For our discussion today we have chosen the sixth grade students and their schools as the major level of focus. At this grade level the dropout rate is not as severe for many minority group students as it is at the higher grade levels and the number of schools in the sample is quite substantial. However, at the higher grade levels the student indexes are more comprehensively measured and the errors in racial-ethnic group identification are less severe than at the lower grade levels hence, some results for these grade levels will also be brought into the discussion. Adequate measures of student attitudes and motivation were not available for grades 1 and 3 and hence these grade levels will not enter further into the discussion.

A MEASURE OF RACIAL-ETHNIC GROUP MEMBERSHIP

We wanted to incorporate in our analyses a variable that would indicate a student's membership in each of the racial-ethnic groups so that we could see how these different groups stood with respect to one another at different points in the analyses. Since these are discrete groups we had to scale or order them in some manner so that a quantitative variable denoting group membership could be incorporated into the analyses. Our primary dependent variable of interest was achievement and consequently we decided to order the groups according to their mean scores on our achievement composite (ACHV). An explanation of this procedure is given in table 1.

Table 1 gives the percent of students[7] in each of the racial-ethnic

[7] In order to obtain more reliable estimates minority group students were oversampled. The percentages in table 1 are weighted by sampling ratios to more nearly reflect population values. Almost half of the students in the sample were from minority groups (Coleman, et al., 1966).

groups along with the mean ACHV score attained by students who identified themselves as belonging to one of these groups. On a distribution with a mean of 50 and standard deviation of 10 we can see that whites attain the highest score with orientals following them by about 4 points. Approximately 5 to 7 points below them lie the Indians, Mexicans, and Negroes with the Puerto Ricans following these groups by another 4 points.

TABLE 1.—PERCENT OF 6TH GRADE STUDENTS AND THEIR AVERAGE COMPOSITE ACHIEVEMENT SCORE CLASSIFIED BY RACIAL-ETHNIC GROUP MEMBERSHIP

Category	Racial-ethnic group	Percent	Mean achievement
1	American Indian	2.6	44.194
2	Mexican-American	6.1	42.244
3	Puerto Rican	2.0	38.560
4	Negro	15.7	42.513
5	Oriental	.9	49.391
6	White	69.6	53.181
7	Other	1.4	45.605
8	No response	1.7	43.144
	Total	100.0	[1] 50.000

[1] The total number of students is 123,386. The standard deviation for the total was equal to 10. All figures are weighted for sampling. Later analyses exclude categories 7 and 8.

Now when each student is assigned the mean ACHV score attained by members of his racial-ethnic group the ordering of these groups is said to be criterion scaled (Beaton, 1969). This means that the relationship of our racial-ethnic group membership variable with ACHV is the maximum relationship that can be obtained. No other ordering of these groups will yield a higher relationship. When scores on our racial-ethnic group membership variable, which we shall call RETH from hereon, are correlated with scores on our ACHV composite the correlation obtained will be a maximum. We are particularly interested in what this maximum value might be and how it changes as different social conditions in which these groups are found are first taken into account.

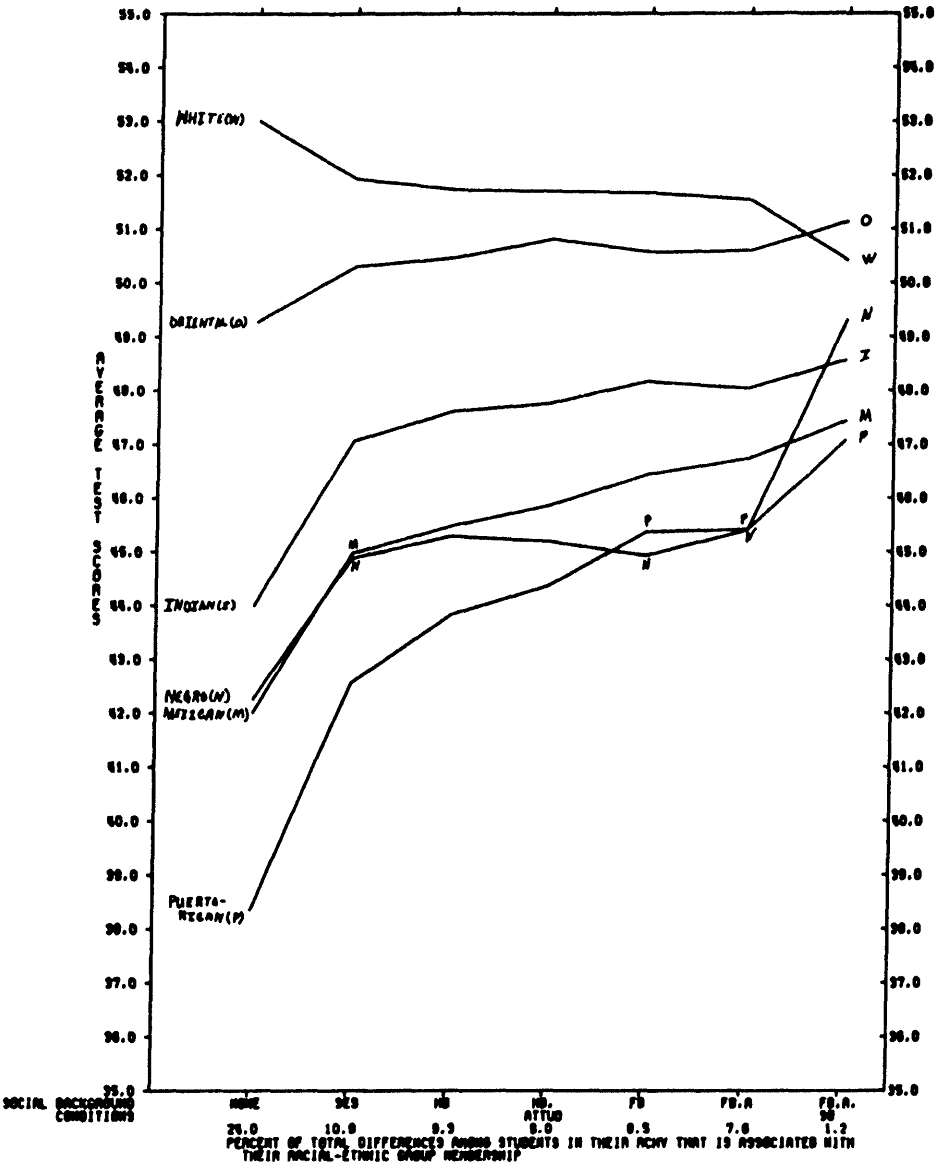
RACIAL-ETHNIC GROUP DIFFERENCES IN ACHIEVEMENT ADJUSTED FOR SOCIAL BACKGROUND CONDITIONS

Our first question then is: "What is the magnitude of this maximum value?" This is indicated by the squared correlation of RETH with ACHV, which is 24 percent and corresponds to the mean differences in the NONE column of figure 1.[8] This is called none because none of the background conditions on which these groups differ has yet been taken into account.

Next we may ask, "What is the percent of variation in ACHV associated with RETH after differences among students in their socioeconomic status (SES) have been taken into account?" A student with a high score on the SES index has parents who come from the higher educational strata, his father is engaged in a professional, managerial, sales or technical job, there are two to three children in the family, about six to 10 rooms in their home, they are more

[8] These mean values differ slightly from those in table 1 due to restandization after the exclusion of students in categories 7 and 8.

FIGURE 1. - RACIAL ETHNIC GROUP ACHIEVEMENT MEANS ADJUSTED FOR SOCIAL BACKGROUND CONDITIONS



likely to reside in the residential area of the city or the suburbs rather than in the inner city and there are intellectually stimulating materials accessible in the home such as books, magazines, newspapers, television, and radio. By taking SES into account we mean that we obtain a squared multiple correlation for both SES and RETH with ACHV when they are entered into the regression together and subtract from this the squared correlation for SES. The resulting value might be more familiar to many if it is called the "unique contribution" of RETH to ACHV.[9] The percent of ACHV associated with RETH after SES has been taken into account is 10.9 percent which corre-

[9] The computational formula is: $U(RETH) = R^2(SES, RETH) - R^2(SES)$. This same computational formula is used throughout, namely, viz., U (RETH) is computed by: $R^2(S, RETH) - R^2(S)$ where S represents the set of variables to be taken into account.

sponds to the differences among the group means in the SES column of figure 1.[10]

The percent of variation in ACHV associated with RETH after other conditions have been taken into account are:

HB—These are the mean differences after considerations of both SES and the students Family Structure (FSS) have been taken into account. They account for 9.3 percent of the differences in ACHV that remain.

HB, ATTUD—These are the magnitude of the mean differences after considerations of SES, FSS and the students Attitude Towards Life (ATTUD) have been taken into account. They account for 9 percent of the differences in ACHV that remain.

FB—These are the magnitude of the mean differences after the indices which we felt represented all aspects of the students Family Background (FB) had been taken into account. These indices were SEC, FSS and the set of four attitudinal and motivational indices (that is, indices 1 through 6 in the appendix). These mean differences account for 8.5 percent of the differences in ACHV associated with RETH that remains.

FB, A—After FB and area of residence (A) whether it be South, Far West or North, or rural suburban or urban have been taken into account, only 7.6 percent remains.

FB, A, SO—After FB. A and the five school attributes of the achievement and motivational mix of the students one goes to school with have been taken into account, only 1.2 percent remains. This set of five student body variables [11] represents a number of things. By virtue of its high correlations with the comprehensive set of 31 school variables,[12] it represents the aggregate effects of schooling. By virtue of its high correlations with the social background of the student body, as defined by their socioeconomic, family structure and racial-ethnic composition,[12] it represents a measure of school and residential segregation.[13]

The trend we observe from figure 1 is that the differences among the racial-ethnic groups in their ACHV levels approaches zero as more and more considerations related to differences in their social conditions are taken into account. This trend is slightly more pronounced for Orientals, whites and Negroes than for the other groups and might be more pronounced for them if more variables pertinent to their special circumstances were also available. We tried English as opposed to some other language spoken in the home but that did not yield any additional information.

GRADE LEVEL AND REGIONAL VARIATIONS

We may next ask: "How do these results compare with those from the other grade levels?" For each of the three grade levels the percent

[10] These means are computed by subtracting from the observed ACHV mean of each group an estimated ACHV mean (using the regression of ACHV on SES) and plotting these as deviations from a mean of 50. This same computational procedure for adjusted means is used throughout (viz., S is used to obtain an estimated mean which is then subtracted from the observed mean and these differences are plotted as deviations from 50, where S represents the set to be taken into account).

[11] There are 2370 schools represented in these analyses.

[12] See the Appendix for these correlations.

[13] Schools being organized according to residential boundaries, circa 1965.

of variation in ACHV associated with RETH before any of the background conditions are taken into account (the NONE condition) and after the FB, A, SO conditions have been accounted for, as these were described in conjunction with figure 1 are:

Grade	None	FB, A, SO
12	20	1.1
9	22	1.0
6	24	1.2

Note: The numbers of students (N) and schools (n) included in these analyses for grades 9 and 12 are, respectively: 9th, N=133,136, n=923, 12th, N=96,426, n=780. Comparable values of RETH with ACHV for the NONE condition are, for the 1st and 3d grades respectively, 16 and 17 percent. These latter values are based on N's and n's of: 1st, N=74 201, n=1,302; 3d, N=130,213, n=2,453.

Thus, for each grade level although the percent before any conditions have been taken into account varies from 24 at the sixth through 22 at the ninth to 20 at the 12th, they all end up at about the same value, namely 1 percent, after a variety of social background conditions (FB, A, SO) have been taken into account. Comparable values for each region [14] are:

Region [1]	6		9		12	
	None	FB, SO	None	FB, SO	None	FB, SO
North nonmetropolitan	17	1	14	0.9	5	0.6
North metropolitan	21	1	20	.9	16	1.8
South metropolitan	26	1	25	.8	28	.1
South nonmetropolitan	30	1	27	.5	30	.9

[1] See the appendix for the States included in the North and South. The metropolitan and nonmetropolitan areas are the census tract standard metropolitan and nonmetropolitan statistical areas (SMSA's and NSMSA's) in each region.

Thus, although the percent of variation in ACHV associated with RETH varies considerably by region, from a high of 30 percent in the nonmetropolitan south to a low of 5 percent in the nonmetropolitan north, they all end up at roughly the same value after adjustments have been made for a variety of social background conditions (FB, SO). Hence, for all practical purposes, all of the differences among students in their academic achievement (ACHV) that are associated with their racial-ethnic group membership (RETH) can be explained by factors that are primarily social in nature and origin.

But, we may ask, wouldn't the slopes of our curves in figure 1 be quite different if we had entered the variables into the regression analysis in a different order? Indeed, this would have been the case, however, it would not have affected our beginning and ending values. We shall see in a moment that we can handle quite well the order of inclusion problem. We might also ask of the relative explanatory power of our variable called RETH when put in context with these other sets of variables. It so happens that our technique for handling the order of inclusion problem also allows us to show the manner in which RETH is inseparably intertwined with these other sets of variable as they relate to ACHV.

---

[14] The set of Area of Residence (A) variables was eliminated from these analyses.

COMMONALITY ANALYSES OF SOCIAL CONDITION VARIABLES AND RACIAL-ETHNIC GROUP MEMBERSHIP WITH ACHIEVEMENT

The technique which allows us to resolve the order of inclusion problem is called commonality analysis. Its introduction at this point increases the order of complexity of the discussion substantially. In an attempt to keep things as simple as possible I shall try to give the essence of this technique for two sets of variables and then move on to a larger number of sets which we really need to address the questions we have posed.

Let us suppose we have two sets of variables. One set shall contain variables that pertain to the structural aspects of a family's position in society which we shall call home background (HB). The other set will be our variable called RETH. Now, following our earlier computational logic, we can compute a "unique contribution" to ACHV for HB and one for RETH as follows:

(1) $U(HB) = R^2(HB,RETH) - R^2(RETH)$

(2) $U(RETH) = R^2(HB,RETH) - R^2(HB)$

The variance that is common to these two sets of variables, or inseparably intertwined among them, can be obtained by the following formula:

(3) $C(HB,RETH) = R^2(HB,RETH) - U(HB) - U(RETH)$

These three coefficients can be organized in tabular form as follows:

	HB	RETH
U(Xi)	a	b
C(X1X2)	c	c
R^2(Xi)	a+c	b+c
R^2(X1X2)	a+b+c	

In this table the row labeled U(Xi) contains the unique coefficients as we shall call them for HB and RETH, denoted as "a" and "b," respectively. Their common portion, or second order commonality coefficient as we shall call it since it involves two sets of variables, is denoted by "c." The squared multiple correlation for each set is given in the row labeled R^2(Xi); for HB it is "a+c" and for RETH it is "b+c."

The squared multiple correlation for both sets is then given by "a+b+c" in the fourth row. This table then contains a complete partitioning of the variance in ACHV explained by these two sets of variables into their common and unique portions.

For those unfamiliar with this technique it may be helpful to think of a Venn diagram in which the area included in two circles represents the variance explained by both sets of variables, the field or background represents the variance unexplained (i.e., $1-R^2(X1X2)$), the overlap of the circles represents their commonality coefficient (C(X1X2)) while the unique contribution of each set (NXi)) is represented by the areas of the circles that do not overlap.

Let us now move to the case of five sets of variables which we shall need to answer the questions we have posed. For the five set case there

will be 2^5-1 or 31 different coefficients that we will organize in tabular form.[15] In the example of our Venn diagram we can think of the different two-way, three-way, four-way and five-way intersections that five circles can have.

The five sets of variables we shall use are:

Home background (HB).—This set includes the indices of socio-economic status (SES) and family structure (FSS). FSS is included here because other analyses have shown it to be highly dependent upon SES in its relationship with ACHV (Mayeske, et al., 1971).

Family process (PRCS).—These are the four attitudinal and motivational indices as described in the appendix (indices 3 through 6). As a set they are called family process because they refer to the expectations and aspirations that a child and his parents have for his schooling and the activities that they engage in to support these aspirations.

Area of residence (AREA).—These are the two variables pertaining to region of residence (North, Far West, South) and rural-suburban-urban location.

Aggregate school outcomes (SO).—These are the five student body variables of the achievement and motivational levels of the students one goes to school with. These are the school averages of the individual student ACHV and PRCS measures.

Racial-ethnic group membership (RETH).—Our racial-ethnic group membership variable as developed and described earlier.

Commonality analyses using these five sets of variables are presented in table 2. From the bottom row of table 2 we can note that 51 percent of the total differences among students in their ACHV can be associated with these five sets of variables. The next to the last row in this table gives the squared multiple correlation for each of the five sets. We can note that this varies from a high of 32 percent for SO through values for HB, PRCS and RETH of 27, 26 and 24 percent respectively, to a low of 4 percent for AREA. Inspection of the unique values in the row labeled U(Xi), shows that the set called PRCS has the largest independent "contribution" with a value of 8 percent. In descending order these unique values for the other sets are 6 percent for SO, 2 percent for HB, 1 percent for RETH and zero for AREA. Hence, for each set of variables most of its variance associated with ACHV is confounded with the other sets of variables. This is completely so for AREA and almost completely so for RETH and HB.

It may be extremely instructive then to look at how these different sets of variables are intertwined or confounded with one another in their relationship with ACHV. Let us first look at the column that contains RETH since its explanatory power has been of major interest to us throughout this paper. RETH can be potentially involved in any of the coefficients (i.e., C()'s) that have the number 5 as a subscript. Of the four second order coefficients involving RETH, only one has a nonzero value (C(X4X5)) and that is the 6 percent in the overlap of RETH with SO. Aside from this 6 percent most of the con-

[15] See Wisler, C. E. "On partitioning the Explained Variation in a Regression Analysis" (Mayeske, et al., 1969).

founding for RETH occurs either in its combination with two other sets of variables or in its combination with three other sets. For example, 4 percent of RETH is confounded with HB and SO while another 7 percent is confounded with HB, PRCS and SO. Clearly then, we cannot make any generalizations about the "independent effect" of membership in a particular racial-ethnic group on ACHV, for this membership is almost completely confounded with a number of social conditions.[16]

TABLE 2.—COMMONALITY ANALYSES OF FAMILY BACKGROUND, AREA OF RESIDENCE, STUDENT BODY ACHIEVEMENT WITH MOTIVATIONAL LEVELS AND RACIAL-ETHNIC GROUP MEMBERSHIP WITH ACHIEVEMENT

	(1) HB	(2) PRCS	(3) AREA	(4) SO	(5) RETH
U(Xi)	2	8	0	6	1
C(X1X2)	5	5			
C(X1X3)	0		0		
C(X1X4)	2			2	
C(X1X5)	0				0
C(X2X3)		0	0		
C(X2X4)		0		0	
C(X2X5)		0			0
C(X3X4)			1	1	
C(X3X5)			0		0
C(X4X5)				6	6
C(X1X2X3)	0	0	0		
C(X1X2X4)	2	2		2	
C(X1X2X5)	1	1			1
C(X1X3X4)	0		0	0	
C(X1X3X5)	0		0		0
C(X1X4X5)	4			4	4
C(X2X3X4)		0	0	0	
C(X2X3X5)		0	0		0
C(X2X4X5)		0		0	0
C(X3X4X5)			1	1	1
C(X1X2X3X4)	1	1	1	1	
C(X1X2X3X5)	0	0	0		0
C(X1X2X4X5)	7	7		7	7
C(X1X3X4X5)	1		1	1	1
C(X2X3X4X5)		0	0	0	0
C(X1X2X3X4X5)	1	1	1	1	1
R SQ (Xi)[1]	27	26	4	32	24
R SQ (total)		51			

[1] Figures do not sum exactly due to rounding.

We noted earlier that of these five sets of variables, RETH was the fourth lowest in its explanatory power. We may ask then of the role played in ACHV by these other sets of variables after considerations of RETH have been set aside. Of the total variation in ACHV associated with these sets of variables about 27 percent (obtained by subtracting 24 from 51) is unrelated to RETH. The variables that play the greatest role in this remaining variation can be ascertained by examining coefficients that do not have a 5 as one of their subscripts (i.e. C(X1X2X3X4)). Most of this variation (i.e., 21 percent) can be explained by the PRCS and HB sets taken together (i.e., the sum of their unique and common portions).[17] The remaining 6 percent is accounted for by the set of school variables (SO). Consequently, some of the most salient variables in explaining differences among students

[16] The nature of this confounding at the school level has been studied extensively in Mayeske, et al., 1969.

[17] Elsewhere we regarded these sets taken together, as representing the student's family background (Mayeske, et al., 1971).

in their ACHV, both before and after considerations as to their racial-ethnic group membership (RETH) have been set aside, relate to the motivational (i.e., PRCS) and social structural (i.e., HB) aspects of the family and to the achievement and motivational levels of the students one goes to school with (SO).

At the higher grade levels similar results were obtained, however, the role [18] played by RETH was somewhat smaller as noted earlier, while the unique "contribution" of PRCS was about twice as large and that of HB and SO about one-half and two-thirds as large respectively as they were at the sixth grade. These results reinforce a point made by Brookover (1963) that we should ask questions not only about class membership but also about attitudes concerned with the value of education and the ways in which these values are operationalized in the parent-child relationship for as we have seen they have a substantial role in ACHV that is independent of social class and racial-ethnic group membership. We should also be humbled by our ability to explain individual academic achievement for almost half of it remains unexplained even though we have experimented with a very wide range of variables (Mayeske, et al., 1971).

SUMMARY

This paper has shown that for sixth grade students, 24 percent of the total differences among students in their academic achievement is the maximum national value that can be associated with their membership in one of six racial-ethnic groups—Indian, Mexican, Puerto Rican, Negro, Oriental or white. This relationship prevails before the allocation of these groups to different social conditions has been taken into account. After a variety of social condition variables have been accounted for such as the social and economic well being of the family, the presence or absence of key family members, the students and parents aspirations for his schooling, their beliefs about how he might benefit from an education, the activities that they engaged in to support these aspirations, one's region of residence and the achievement and motivational levels of the students one goes to school with, this percentage dropped to 1.2. Similar results were obtained for other grade levels and for each region of the country. Hence, no inferences can be made about the "independent effect" of membership in a particular racial-ethnic group on academic achievement, for that membership, as it relates to academic achievement, is almost completely confounded with a variety of social conditions.

Other analyses showed that variables pertaining to the motivational and attitudinal aspects of family life—what one might term "educationally related child rearing activities"—played a greater independent role in academic achivement than did either racial-ethnic group membership, social class membership, or the type of school one attended. Of these latter two, however, the type of school one attended had a somewhat greater independent role than did one's social class membership.

---

[18] By role we mean the variance in ACHV associated with RETH before and after adjustments had been made for FB, A, SO.

LIST OF REFERENCES

Beaton, A. E., Jr., "Criterion scaling of questionnaire items for regression analysis." Socio-Economic Planning Sciences, 1969, (2), 335–362.

Brookover, W. and Gottlieb, D., "Social Class and Family Influences" in Charters and Gage (Eds.) "Readings in the Social Psychology of Education." Boston, Mass.: Allyn and Bacon, Inc., 1963.

Coleman, J. S., et al., "Equality of Educational Opportunity." Washington, D.C.: U.S. Government Printing Office, 1966.

Mayeske, G. W. and Weinfeld, F. D., "Factor analysis of achievement measures from the Educational Opportunities Survey." Technical Note No. 21, U.S. Office of Education, Washington, D.C., 1967.

Mayeske, G. W.; Weinfeld, F. D. and; Beaton, A. E., Jr. "Item response analyses of the Educational Opportunities Survey teacher questionnaire." Technical Note No. 32, U.S. Office of Education, Washington, D.C., 1967.

Mayeske, G. W.; Tabler, K. A.; Proshek, J. M.; Weinfeld F. D. and; Beaton, A. E., Jr. "Item response analyses of the Educational Opportunities Survey principal questionnaire." Technical Note No. 58, U.S. Office of Education, Washington, D.C., 1968.

Mayeske, G. W.; Weinfeld, F. D.; Beaton, A. E., Jr.; Davis, W.; Fetters, W. B.; Hixson, E. E., "Item response analyses of the Educational Opportunities Survey student questionnaire" Technical Note No. 64, U.S. Office of Education, Washington, D.C., 1968.

Mayeske, G. W., et al., "A study of Our Nation's Schools." Washington, D.C., U.S. Government Printing Office, 1969 (A Working Paper).

Mayeske, G. W., et al., "A Study of the Achievement of Our Nation's Students." Washington, D.C., U.S. Government Printing Office, 1971 (In Press).

APPENDIX

INDIVIDUAL STUDENT INDICES AND VARIABLES

Socioeconomic status (SES): A student with a high score on this index has parents who come from the higher educational strata, his father is engaged in a professional, managerial, sales or technical job, there are two to three children in the family, about six to 10 rooms in their home, they are more likely to reside in the residential area of the city or the suburbs rather than in the inner city and there are intellectually stimulating materials accessible in the home such as books, magazines, newspapers, television, and radio.

Family structure and stability (FSS): A student with a high score on this index has both parents in the home, his father's earnings are the major source of income, his mother works part time or not at all and his family has not moved around much.

Racial-ethnic group membership (RETH): A student with a high score on their variable is white, a student with an intermediate score is Oriental-American and a student with a low score is Puerto Rican, Mexican-American, Indian-American, or Negro-American.

Expectations for excellence (EXPTN): A student with a high score on this index says that his mother, father, and teachers want him to be one of the best students in his class and that he also desires to be one of the best in his class.

Attitude toward life (ATTUD): A student with a high score on this index feels that: People who accept their condition in life are not necessarily happier; hard work is more important than good luck for success; when he tries to get ahead he doesn't encounter many obstacles; with a good education he won't have difficulty getting a job; he would not sacrifice anything to get ahead nor does he want to change himself; he does not have difficulty learning nor does he feel that he would do better if his teachers went slower; and, people like him have a chance to be successful.

Educational plans and desires (EDPLN): A student with a high score on this index says that his parents want him to go to college, he both desires and plans to go to college, aspires to one of the higher occupational levels and feels that he is one of the brighter students in his class.

Study habits (HBTS): A student with a high score on this index has frequent—weekly or more—discussions with his parents about his school work, was read to regularly as a child, spends 1 to 3 hours a day studying and 1 to 3 hours a day watching TV, would make most any sacrifice to stay in school and has seldom stayed away from school just because he wanted to.

Achievement (ACHV): A student with a high score on this index or composite tended to score high on all of the tests that entered into that composite. For all grade levels the tests of verbal and nonverbal ability were used as part of the composite. In addition, at grades 6, 9, and 12, tests of reading comprehension and mathematics achievement were used and at grade 9 and 12 a test of general information was in-

cluded in the composite. In one sense, this inclusion of more tests at the higher grade levels represents the nature of the educational process, where basic skills are acquired in the early years and then other skills and knowledge are acquired through the use of these basic skills.

The standard deviations of this composite for each group at the different grade levels are:

Group	Grade level 1	3	6	9	12
Indian	1.70	2.23	2.84	3.20	3.65
Mexican	2.03	2.14	2.76	3.45	3.42
Puerto-Rican	2.26	2.33	2.58	3.34	3.72
Negro	2.04	2.11	2.38	3.09	3.30
Oriental	1.38	2.69	3.35	3.61	3.84
White	1.56	2.15	2.65	3.27	3.22
Total	1.87	2.37	3.01	3.67	3.63

Student Body Variables

When the values of a variable are averaged for each of the students in a particular grade level of a school, this results in what is called a student body variable. Schools with a high mean or average on a student body variable tend to have a larger proportion of students with a high score on that attribute, while schools with a low mean or average tend to have a larger proportion of students with a low score on the attribute. The meaning of these variables at the individual student level were indicated earlier. The student body variables used in this paper are:

- Student body expectations for excellence.
- Student body attitude toward life.
- Student body educational plans and desires.
- Student body study habits.
- Student body achievement.

School Variables

In these studies, to represent attributes of the schools other than student body variables, the following comprehensive set of 31 indices and variables was used. A description of the meaning of each index and the variables that comprise it is given in Mayeske, and others, 1971. A detailed description of the development of these indices and variables is given in Mayeske, and others, 1969. The indices and variables are categorized into three subsets of facilities, pupil programs and policies, and school personnel expenditures. All but seven of the 31 are indices.

FACILITIES

(1) Plant and physical facilities
(2) Instructional facilities
(3) Pupils per room
(4) Age of buildings

PUPIL PROGRAMS AND POLICIES

(1) Tracking
(2) Testing
(3) Transfers
(4) Remedial programs
(5) Free milk and lunch programs
(6) Accreditation
(7) Age of texts
(8) Availabiilty of texts
(9) Pupil-teacher ratio
(10) Enrollment

SCHOOL PERSONNEL AND PERSONNEL EXPENDITURES

(1) Principal's experience
(2) Principal's training
(3) Principal's college attended
(4) Principal's sex
(5) Principal's estimate of the school's reputation
(6) Specialized staff and services
(7) Teacher's experience
(8) Teacher's training
(9) Teacher's socio-economic background
(10) Teacher's localism
(11) Teacher's college attended
(12) Teaching conditions
(13) Teaching related activities
(14) Preference for student ability level
(15) Teacher's sex
(16) Teacher's racial-ethnic group membership
(17) Teacher's vocabulary score

In the accompanying paper the above set of variables is referred to as the comprehensive set of 31 school variables.

At the sixth grade the squared multiple correlations of the comprehensive set of 31 school variables (SCH) and the set of 3 student body social background (SBSB)[1] variables with each of the five school outcome (SO) variables are:

	Expectations for excellence	Attitude toward life	Educational plans and desires	Study habits	Achievement
SCH	23	30	33	28	73
SBSB	46	57	50	57	79

DEFINITION OF GEOGRAPHIC GROUPINGS

For the quantiative variable called "area of residence" the following scales were used:

Rural-Suburban-Urban.—A seven-point scale with a low value for rural-through-intermediate values for small towns and cities to large values for suburbs, residential, and inner parts of large cities.

Region.—A three-point scale, scored: low for South (Alabama, Arkansas, Arizona, Florida, Georgia, Kentucky, Louisiana, Mississippi, New Mexico, North Carolina, Oklahoma, South Carolina, Texas, Tennessee, Virginia, and West Virginia); intermediate for the Far West and Rocky Mountain States of Alaska, California, Colorado, Hawaii, Idaho, Montana, Nevada, Oregon, Utah, Washington, and Wyoming; high, for the 23 Northern States.

When regional stratifications were used, included as south were the above 16 southern States; all other States were included as north.

[1] The SBSB set is comprised of the three student body variables of: Socio-Economic composition; Family Structure; and Racial-Ethnic Composition.

DYSGENICS—A SOCIAL-PROBLEM REALITY EVADED BY THE ILLUSION OF INFINITE PLASTICITY OF HUMAN INTELLIGENCE?*

By William Shockley

1. RESOLVING THE ENVIRONMENT-HEREDITY UNCERTAINTY

My chief contribution to this symposium is to ask a question—an unpleasant question, but one that I believe must not only be asked but answered if our generation of citizens is to fulfill its responsibility to the next generation. My question is:

> Do important social problems arise from dysgenics—retrogressive evolution through the disproportionate reproduction of the genetically disadvantaged?

Underlying this question is the nature-nurture issue. I described it in 1966 as the environment-heredity uncertainty in order to draw parallels with those uncertainty principles in physics that are basic. My thesis today is that the environment-heredity uncertainty is not basic and indeed it has really been resolved—at least for one significant case that I shall discuss—but that an illusion or a delusion prevents the acceptance of the reality of this resolution and blocks its application to the social problems being faced by this symposium.

The resolution of the environment-heredity uncertainty that I shall describe is limited to the IQs of individuals in one particular population; and further acceptable research is needed for a comparable resolution applicable to social problems for the U.S. population as a whole.

Because these limitations prevent evaluation of the dysgenic threat, I have demanded increased research on genetic aspects of human-quality problems. Four of the most frequent reasons given for rejection of my demands are these: (1) intelligence measured by IQ score is so complexly influenced by culture that genetic influences are not quantifiable; (2) IQ score has no relevance to successful living; (3) races cannot be meaningfully defined and all ethnic groups have the same genetic potential for intelligence; and (4) even if the environment-heredity uncertainty, including its racial aspects, were resolved, the knowledge would be worthless because the needed remedies would inevitably require quality control applied to human reproduction on the basis of genetics. This is nothing less than eugenics—a repugnant concept.

As I shall demonstrate in the remainder of my presentation, none of these four objections stand up under objective analysis.

---

*Manuscript planned for reading at the American Psychological Association, Division 9, Society of Psychological Study of Social Issues, Symposium on "Social Problems: Illusion, Delusion of Reality." Washington, D.C., Tuesday, September 7, 1971.

77-942 O—72——36

2. GENETICITY OF IQ AND THE SIGNIFICANCE OF THE GLADYS-HELEN CASE

Slide 1 is my answer to the first objection. I use published data to "predict" 122 "observed" IQs. The root-mean-square error of prediction is only 8.5 IQ points for the 122 cases that are distributed with a standard deviation of 15 points. The "prediction" is possible because four studies have matched each "observed" IQ with the IQ of an identical twin reared apart. This other IQ is my "prediction;" each point is a twin pair. I maintain, but most psychologists deny, that the details of these studies assembled by A. R. Jensen from England, Denmark, and the United States validate this assertion:

> Intelligence, measured by IQ, varies more than twice as much from genetic difference as from environmental ones for individuals from families like those that raise one of a pair of white identical twins. This assertion is conservative. The correlation coefficient between twins' IQs is 0.82: "geneticity" [i.e., my nondictionary word, like "culturology" of this symposium, for the fraction of population variance due to genes] is 82 percent; nongenetic factors cause only 18 percent of the variance.

If the results of slide 1 are as obvious as I assert, why are they not accepted?

The twin data of slide 1 can be differently—but not soundly— interpreted. In fact, one pair of twins in the study of Newmann, Freeman, and Holzinger have been repeatedly cited as evidence for what I label the *illusion of infinite plasticity of intelligence.* Gladys and Helen differed by 24 IQ points—much more than the average IQ difference between whites and Negroes. Obviously, it is asserted, environment has dominant control.

This reasoning, that is emphasized in many psychology texts, is superficial. Actually the Gladys-Helen case provides an exception needed to prove the 82-percent geneticity rule. Failure to interpret these results soundly seems to me an example of the myths about social problems that this symposium may dispel.

The correct reasoning is presented in slide 2. In brief, what it shows is that nongenetic contributions to IQ differences between the twins are accurately distributed in a normal distribution. One striking result on this slide is that the famous pair of identical twins, Gladys and Helen of the well-known Newmann, Freeman, and Holzinger study, do indeed supply the exception that proves the rule. In a distribution of 122 pairs of twins, *one pair differing by 24 IQ points should be found by the laws of probability if geneticity is 82 percent.*

The normal distribution of slide 2 also warrants another important conclusion—one not previously presented at a scientific meeting so far as I know. This new conclusion is an evaluation of the confidence that one can place in the 82-percent geneticity value—always, of course, for populations like those that raise one of a pair of white identical twins. My own research on this older research reveals that *if all the nongenetic factors that affected the IQs added up to as much as 29 percent of the total variance, then there is less than one chance in 2,000 that chance alone would have produced the smallness of the observed 112 IQ dif-*

ferences between the separately reared co-twins. In other words, *the greater importance of genes compared to environment is established at a level of significance enormously higher than one in 2,000.* Geneticity is most unlikely to be less than twice as important as everything else—always for the limited population considered. Further research shows that this conclusion is not a spurious consequence of similar environments for both twins of a pair.

One prediction from 82-percent geneticity is that a difference of approximately 25 IQ points between identical twins should occur if one is raised in the worst 1 percent and the other in the best 1 percent of the normal distribution of environments. This may be relevant to the recently publicized results for young slum children reported by Professor Heber of Wisconsin.

Regarding the second objection—IQ means nothing—I observe that IQ is positively correlated with many socially-accepted measures of human quality. I refer you to A. R. Jensen's well-known article, H. J. Eysenck's recent book and Richard Herrstein's article in the current Atlantic Monthly for data on traits that I calculate have correlation coefficients of about 0.2 to 0.5 with IQ.

3. RACEOLOGY AND THE MORAL OBLIGATION TO DIAGNOSE

The third objection—that race is meaningless—is refuted by T. E. Reed of Toronto who has determined with a precision of 1 percent that the Oakland, Calif. Negro population is 22 percent Caucasian in ancestry. I have refined Reed's studies and used them with Army pre-induction test data to estimate that for low IQ Negro populations, each 1 percent of Caucasian ancestry raises average IQ by 1 point. I have suggested ways of controlling for the environmental differences to test the reliability of this estimate. An interesting question is the level at which diminishing returns set in; for example, at 40-percent Caucasian ancestry, would average IQ be 110?

In respect to this symposium's concern with "social problems" and its goal of "the reestablishment of stability, order and meaning," I express this warning: To fail to use diagnosis based on racial differences in blood types for fear of being called a racist is irresponsible. It may also be a great injustice to black Americans themselves. If those Negroes with the fewest Caucasian genes are in fact the most prolific and also the least intelligent, then genetic enslavement will be the destiny of their next generation. The consequences may be extremes of racism and agony for both blacks and whites.

The word "raceology" has been proposed for studies like mine. They are not racism. They are motivated by concern—not by fear and hate. My research focuses principally upon white-Negro comparisons for two reasons: (1) Our national racial problems primarily involve the Negro minority and (2) Negroes are the only racial group for which extensive published statistics are available. Therefore, my personal research on questions related to Negroes has far greater immediate promise of contributing to sound diagnosis of our human quality problems than, for example, would attempts to study hereditary factors for Appalachian whites, for whom I have found that statistical data is practically unobtainable. Although I emphasize the Negro

area for these reasons, I continue to urge broad inquiry into hereditary aspects of human behavior for all racial groups.

As an example of raceology, I present in slide 3 some new research results on Negro superiority that compares Negro and white visual acuity as based on Army tests. The points specify fractions of Negroes and whites having various levels of visual acuity. From 20/20 to less than 20/200, the points fall accurately along a line. The interpretation of this analysis is that whites and Negroes are distributed in their visual acuity according to the same basic underlying normal distribution but that the distribution for Negro visual acuity is offset upwards by approximately 0.6 of a standard deviation—a value that if it applied for mental performance would be equivalent to about 9 IQ points.

Where data have been available, I have tried to compare other racial groups. My findings do not support a theory of white Aryan supremacy: I have found and published the observation that American Orientals are about 10 times more successful than the national average on a per capita basis in achieving the distinction of election to the National Academy of Sciences. They are also about 10 times more successful in avoiding citations in the annual FBI uniform crime reports. My statistics also show that Jewish Nobel Prize winners in science occur about 10 times more often than expected on the basis of the population as a whole.

4. THE "APPLE OF GOD'S EYE OBSESSION"—A CAUSE OF DELUSIONS ABOUT SOCIAL PROBLEMS?

I shall now attempt an analysis of psychological factors underlying the four objections to my research demands. I shall start with the fourth—that knowledge would be worthless because any possible action would involve intolerable eugenics measures.

Eugenics is a shunned word because it was a feature of Hitlerism. But the lesson of Nazi history is not that eugenics is intolerable. Denmark has continued since 1935 programs with clearly positive eugenic implications. One hundred and forty years before Hitler, our Bill of Rights anticipated the lesson to be learned from Nazi history by incorporating into our Constitution the first amendment guaranteeing freedom of speech and of the press. Only the most anti-Teutonic racist can believe that the German people are such an evil breed of man that they would have tolerated the concentration camps and gas chambers if a working first amendment had permitted exposure and discussion of Hitler's "final solution"—the extermination of the Jews.

The first amendment makes it safe for us in the United States to try to find humane eugenic measures. As a step in such search, I propose as a "thinking exercise," a voluntary, sterilization bonus plan.

Bonuses will be offered for sterilization. Income taxpayers get nothing. Bonuses for all others, regardless of sex, race, or welfare status, would depend on best scientific estimates of hereditary factors in disadvantages such as diabetes, epilepsy, heroin addiction, arthritis, et cetera. At a bonus rate of $1,000 for each point below 100 IQ, $30,000 put in trust for a 70 IQ moron of 20-child potential might return $250,000 to taxpayers in reduced costs of mental retardation care. Ten

percent of the bonus in spot cash might put our national talent for entrepreneurship into action.

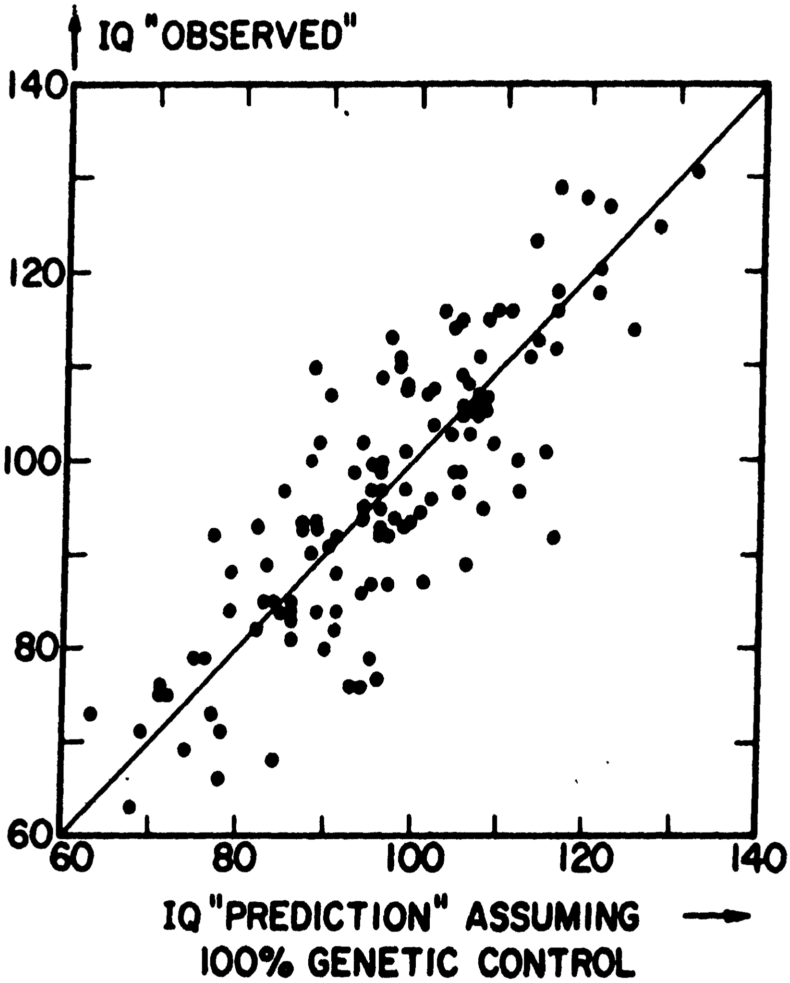
A motivation boost might be to permit those sterilized to be employed at below minimum standard wages without any loss of a welfare floor income. Could this provide opportunity for those now unemployable?

I shall close with an hypothesis about the psychology of the critics of my concerns about dysgenics. I doubt neither the sincerity nor the good intentions of these critics. I diagnose their obtuseness as caused by a theologico-scientific delusion. I call it the "Apple of God's Eye Obsession"—God meaning for some the proper sociobiological order of the universe. True believers of this obsession hold that God has designed nature's laws so that good intentions suffice to insure humanity's well-being—a belief that satisfies a human need for self-esteem. Any evidence counter to man's claim to be the apple of God's eye strikes a central blow at his self-esteem, and thereby provokes retaliation reminiscent of the prompt execution of a Greek messenger bearing ill tidings of defeat in battle. These parallels become clearer in the historical perspective of Galileo and Darwin. In each case they brought new knowledge that was incompatible with the then cherished interpretation of humanity's unique place in the universe. Either the new knowledge had to be rejected or else the "Apple of God's Eye Obsession" had to be painfully revised.

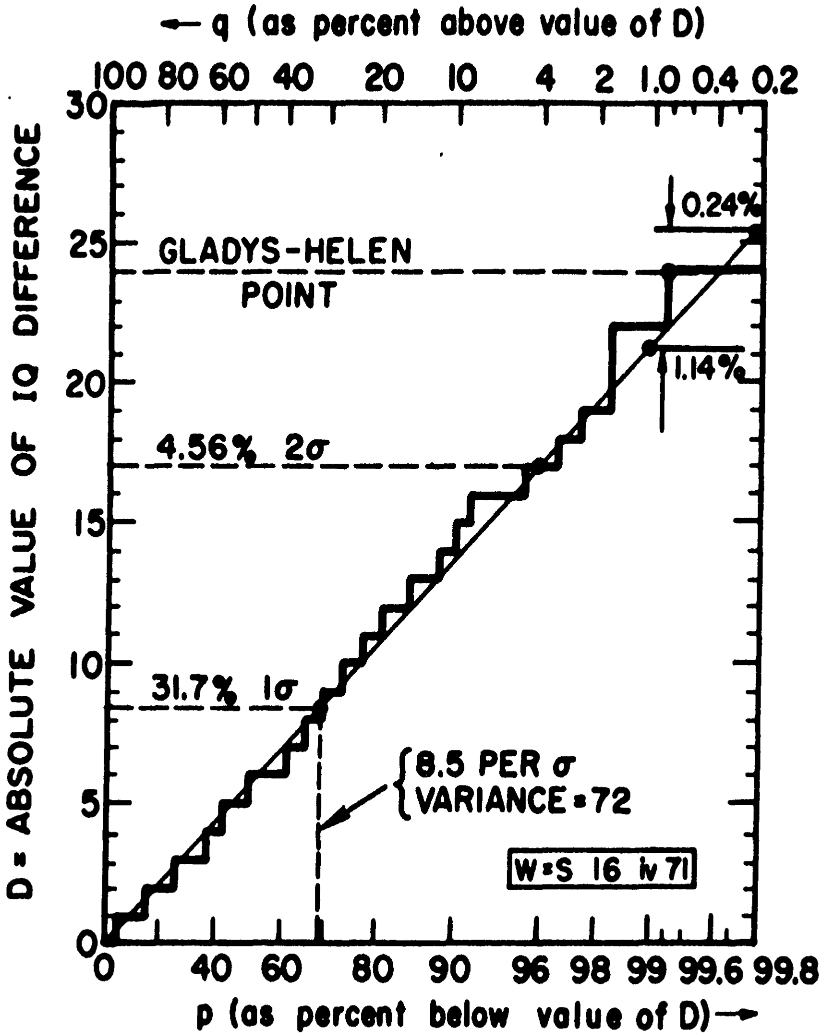
I propose that illusions and delusions are important in the rejection of the relevance of genetics to social problems because the theory that intelligence is largely determined by the genes and that races may differ in distribution of mental capacity offends equalitarian-environmentalism—an important feature of the contemporary form of the "Apple of God's Eye Obsession." The preponderance of the world's intellectual community resists the fact that nature can be cruel to the newborn baby. Babies too often get an unfair shake from a badly loaded parental genetic dice cup. At the acme of unfairness are features of racial differences that my own research inescapably leads me to conclude exist: Nature has color-coded groups of individuals so that statistically reliable predictions of their adaptability to intellectually rewarding and effective lives can easily be made and profitably be used by the pragmatic man-in-the-street.

If, as many thinking citizens fear, our welfare programs are unwittingly, but with the noblest of intentions, selectively down-breeding the poor of our slums by encouraging their least foresighted to be most prolific, the consequences will be tragic for both blacks and whites—but proportionately so much worse for our black minority that, as I have said, the consequence may be a form of genetic enslavement that will provoke extremes of racism with agony for all citizens.

My position is that humanity has an obligation to use its intelligence to diagnose and to predict in order to prevent agonies that lack of foresight can all too easily create. The ambition of this symposium to dispose of "illusions and delusions" by "delving deeply into the social issues of our day" and seeking "solutions * * * which draw from man's basic core: his meaning system * * *" are in keeping with my position. I consider it a privilege to participate.

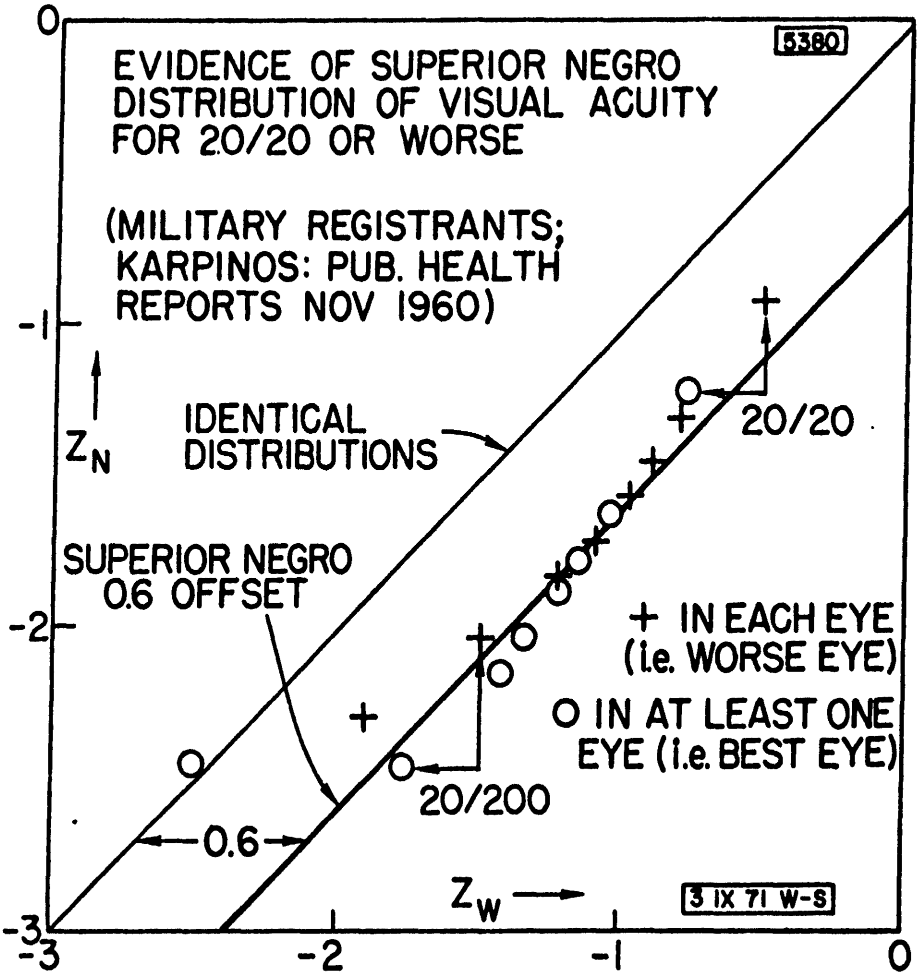


Actually each "prediction" is the IQ of one of a pair of separately-reared, white-identical twins. The "observed" value is the other. The correlation coefficient is 0.82 implying that only 18 percent of the population variance is nongenetic. Thus "geneticity" or fraction of variance due to genetic differences is 82 percent.

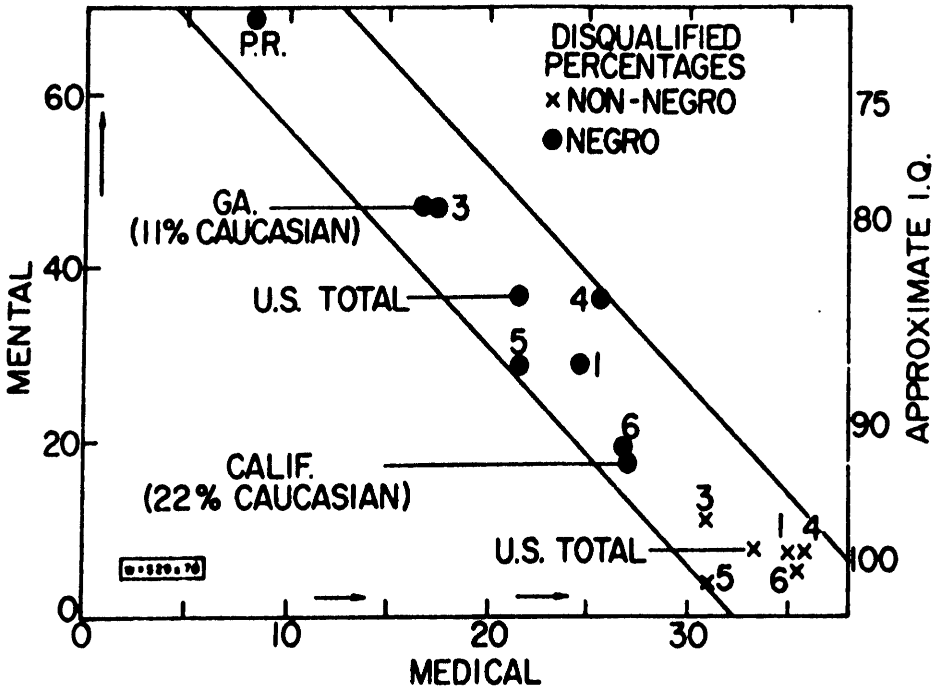


This figure shows that differences in IQ between identical twins reared apart obey a basic statistical law known as the normal distribution. If the data that give the "staircase" of heavy lines fell so that a straight line cut each step in half, the fit would be perfect—in fact, too perfect—like perfect alternation between heads and tails for a tossed coin. The figure shows that Gladys and Helen, the identical twins famous for differing by 24 IQ points, are the exception that proves the rule—the normal distribution predicts one such case among the 122 pairs of twins just as six heads in a row occurs once in 64 tries.

The data of the figure warrant the assertion that *intelligence, measured by IQ, varies more than twice as much from genetic differences as from environmental ones for individuals from families like those that raise one of a pair of white identical twins.* If genetic differences were less than twice as important as environmental ones, the probability is less than one in 2,000 that chance would have produced the good fit of the figure.



The Z_w values give normal distribution arguments that correspond to the percentage of white, military registrants who fail to meet the prescribed visual acuity. Z_N corresponds to Negroes. The unmarked visual acuities are in sequence 20/30, 20/40, 20/50, 20/70, 20/100. The extreme points that fall out of the pattern are 20/400. If the points fell perfectly on the line, it would imply identical normal distributions for both races except for an offset of 0.6 standard deviations.



Evidence that increases in percentages of Caucasian genes in Negro populations improve mental performance and degrade physical performance is furnished by the preinduction test results reported by the Office of the Surgeon General, Department of the Army. The 1968 results show that Negroes in Georgia in the Third Recruiting District have a mental disqualification rate of 47.3 percent or an IQ of about 80 compared to 17.5 percent and 90 for California in the sixth district. The superior performance of Negroes in California compared to Georgia supports the theory that Negro IQ is raised by an admixture of white ancestry. California Negroes have twice as high a percentage of their genes from white ancestors as do Georgia Negroes according to an estimate based on measurements by Prof. T. E. Reed of the University of Toronto of 22 percent Caucasian genes for Oakland, Calif., and 11 percent for Evans and Bullock Counties, Ga. Reasoning from the trend shown by all the recruiting districts for both Negro and non-Negro inductees, Prof. William Shockley estimates that the average IQ of Negro populations increases by about 1 IQ point for each 1 percent of added Caucasian genes and might match or even exceed the whites at 30 or 40 percent. The physical qualifications correspondingly drop. Professor Shockley urges that his hypothesis should be tested by determining the percentages of Caucasian genes for representative populations of Negro inductees. Such research might also permit evaluating the claim that Negro-white differences in medical disqualifications are biased by the poor medical counseling available to the economically disadvantaged.

[From the Phi Delta Kappan, January 1972]

DYSGENICS, GENETICITY, RACEOLOGY—A CHALLENGE TO THE INTELLECTUAL RESPONSIBILTY OF EDUCATORS

By William Shockley

William Shockley is Alexander M. Poniatoff, professor of engineering sciences, Stanford University. In 1956 he was cowinner (with John Bardeen and Walter H. Brattain) of the Nobel Prize in Physics for invention of the transistor. For several years he has pursued an interest in the genetic factors in intelligence. He has urged the National Academy of Sciences to encourage systematic study of the relative influence of heredity and environment on human intelligence and genetic factors in human performance. Although unsuccessful in this effort to date, Shockley continues to call for examination of these and related questions.

In September 1971, Mr. Shockley presented a paper before division 9, Society for Psychological Study of Social Issues, American Psychological Association, at its Washington, D.C., meeting. His topic was, "Dysgenics: A Social-Problem Reality Evaded by the Illusion of Infinite Plasticity of Human Intelligence?" This article is an elaboration of ideas presented in that paper.

Do our nobly intended welfare programs promote dysgenics—retrogressive evolution through the disproportionate reproduction of the genetically disadvantaged? One incident that led me to express my worries publicly was a news story of an acid-throwing teenager, one of 17 children of a mother with an IQ of 55. Later I learned of Denmark's sterilization programs with their eugenic implications. The rising per capita homicide rate of Washington, D.C., is 50 times Denmark's falling one. Dysgenics?

My inquiries unearth no support of studies of dysgenics by a government agency or a major foundation. But conspicuous hints of dysgenic worries do occasionally emerge. In 1964 Secretary of Labor Willard W. Wirtz said: "There is a strong indication that a disproportionate number of unemployed come from large families, but we don't pursue evidence that would permit establishing this as a fact or evaluating its significance." [1] Early in 1971, Vice President Spiro T. Agnew mentioned forbidding welfare mothers to have more illegitimate children and suggested that welfare problems might require

[1] Willard W. Wirtz, OCED speech, 1964, and personal correspondence with the author.

willingness "to take on the hard social judgments that very frankly no one that I know in elective office is willing to even think about."[2] This unwillingness is not restricted to politicians. Agnew's thoughts were rejected with the adjectives "punitive" and "inhumane" in an article in *Science*.[3]

Inverted liberals of our academic community encourage this we-don't-pursue, no-one-willing-even-to-think avoidance of dysgenics by our political leaders. They devise such unsearch dogmatism as this rephrased thought-blocker: "An individual's IQ is controlled by two variables, his environment and his genes. Separate control of these variables is neither practical nor humane. Therefore, to determine the 'geneticity' [my word for the genetic fraction of the spread, precisely of the variance or square of standard deviation] of IQ for any population is impossible. Environmental improvements in human quality so need resources that none should be wasted on 'bad heredity' research."

To refute the unsearch dogmatism of the above "two-variable-basically-impossible" thought-blocker, I exhibit figure 1, showing my use of published research to "predict" 122 "unknown" IQs, together with the "observed" values.

A CHALLENGE TO THE READER

I challenge Kappan readers to answer this question: How can these genetically based "predictions" be possible? This is the question that my audiences ask me when I project figure 1 as a slide. They ask: "Do you use the IQs of the parents?" I reply: "Parents' IQs do not permit such accuracy. The predictions of figure 1 account for 82 percent of the IQ variance of the 'observed' population. There is only one way it can be done."

Dear reader, does a thought-blocker prevent you from recognizing the familiar because I have presented it in an unfamiliar light? These "100 percent genetic control predictions"—I phrase this with scrupulous precision—can be made in only one way—a way that you know if you remember a good psychology course. If you can't dispose of my challenge, is the "Apple of God's Eye Obsession" the cause of your thought-block? Will any of you suffer "Speer syndrome" a decade or two from now? I define these concepts in my conclusion (page 305), "The Moral Obligation to Think."

Associated with my challenge are two questions: (1) On what do I base my "predictions"? (2) How can one sort out the environmental influences quantitatively after one does know the basis? I ask the reader to be my student while I elucidate a pedagogical methodology that permits the necessary analysis of variance to be understood by one whose mathematical skills are at the precollege level. While you read, keep my challenge in mind. Perhaps, before my explanation leaves no challenge to meet, you will overcome the thought-block that most of my audiences experience on encountering figure 1.

---

[2] Earl C. Behrens, "Tough Agnew Proposals on Welfare," *San Francisco Chronicle*, January 15, 1971, pp. 1, 24.

[3] Frederick S. Jaffe, "Toward the Reduction of Unwanted Pregnancy," *Science*, October 8, 1971, pp. 119–27.

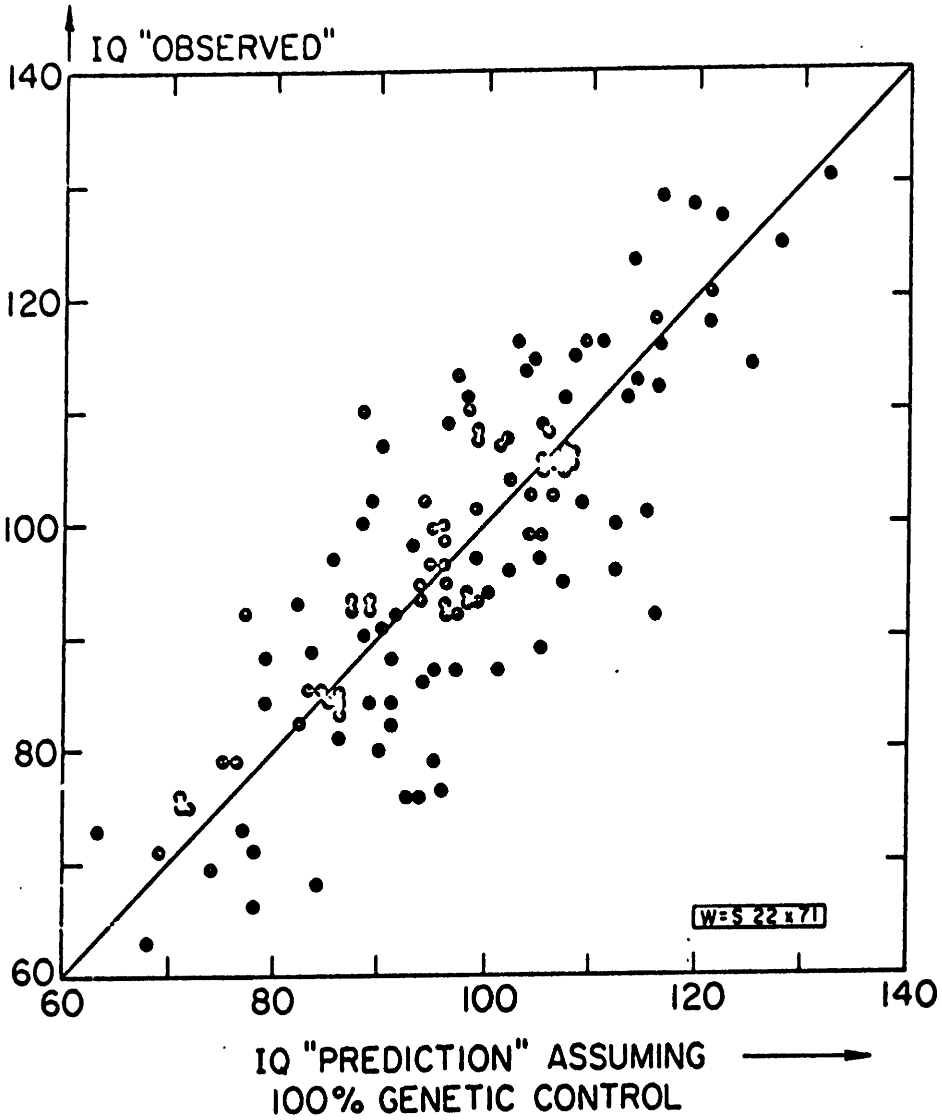


FIGURE 1.—The challenge to *Kappan* readers: How could such accurate predictions of IQ's be made on the basis of the assumption that IQ is 100-percent controlled by the genes?

Now back to figure 1. The average of the 122 "observed" IQs is 96.8 and the standard deviation is 14.2. Furthermore—and this is important in what follows—the distribution is typical of representative Caucasian populations and is accurately normal over the range covered by the 122 cases. The same applies to the "predicted" distribution. For simplicity, we round these off to an average of 100 with a variance of 200 (14.2 squared=201.64).

The "Las Vegas" method, my Americanized version of the Monte Carlo method of statistics, consists of creating a normal distribution generator in the form of a deck of cards from which randomly drawn cards produce a set of positive and negative integers that may represent genetic or environmental contributions to whatever pushes IQ up and down around the population norm of 100. Analysis of variance

then consists simply of finding by trial and error what mix of environmental and genetic influences will duplicate the actual fact of figure 1. The result, which I shall teach you how to duplicate on your own, is shown in figure 2. In part (a), the genetic weight is four times the environmental weight, that is, geneticity is 80 percent; environmental differences contribute only 20 percent of the variance. Part (a) was produced by drawing four genetic cards and one environmental card, all from the same deck, to get each "observed" IQ. It is seen to represent figure 1 very well. In part (b), the ratio is altered to three genetic and two environmental; it is obviously a poor fit; the predictions of figure 1 could not have worked out so well had geneticity been so small as 60 percent.

I shall not at this point of my exposition explain exactly how to apply the card drawing ratios to represent the mysterious prediction process of figure 1; to do so would deprive you of the opportunity to respond to my challenge. After the challenge is disposed of, the procedure for combining the genetic and environmental cards will be obvious. Next I shall explain how to mark 50 cards from an ordinary deck so that a random choice of five will give scores that on the average add to zero and have a variance of 200 and approximate a normal distribution. This is done by marking 50 cards (some felt-tip marking pens are excellent) as follows: Take 25 black cards and mark them with these numbers: 0,0;1,1,1;2,2,2;3,3,3;4,4;5,5;6,6;7,7;8,8;9;10;12;15. Do the same with 25 red cards. Count the black cards as plus and the red as minus—after all, being "in the red" is minus. The symmetry of plus and minus ensures that the average of many draws is zero. Tests will show you that the variance must be 40, because variances add for independent contributions and you will find that five cards do match the 200 variance of figure 1.

To convince yourself that the geneticity of figure 1 is about 80 percent—certainly more than 60 percent—you need not understand the theory of the S-N50-V40 deck—that is, the Shockley Normally distributed 50-card deck with Variance of approximately 40; precisely, 38.9. The point of the method is that random draws of four genetic cards to one environmental cards does indeed match the reality of figure 1. A ratio of three to two fails badly.

What about my challenge? The quotation marks on "observed" and "predicted" have been a broad hint. The next paragraph—Stop! *If you look before you resolve the challenge you become one more item of evidence for the thought-blockage that afflicts our Nation's intellectual community on matters of human genetic quality*—gives the obvious and familiar answer—an answer that typically only 1 or 2 percent of my college audiences can produce when the projection of a slide emphasizes the shocking evidence for the dominance of genetic differences over environmental ones in pushing IQ scores around—especially shocking to the educational fraternity, whose income would burgeon if they could discover how to convert retardates into geniuses.

GENETIC DOMINANCE OF IQ: "LAS VEGAS" ANALYSIS, SIGNIFICANCE LEVEL

A dispassionate appraisal of the existing data (that of figure 1 is the best and the easiest to understand, but the same conclusions can be reached without it) leads to the conclusion that intelligence, meas-

ured by IQ, varies more than twice as much from genetic differences than it does from environmental differences for individuals from families like those that raise one of a pair of white identical twins. The only reason that the conclusion of the preceding sentence is not printed in bold-face in a display paragraph is that it would have given away the answer to my challenge too easily. Did you guess it? Identical twins, reared apart, are the naturally occurring experiment that gets around the "two-variable-basically-impossible" thought-blocker discussed above. The 122 "predictions" of IQ are obtained by reading from one column of a compilation published by A. R. Jensen.[4] If you cover the adjacent column, then the IQs of the other twin will be "unknown" to you. For example, take the highest IQ "predicted" in figure 1: The uncovered column shows 132; the covered column is found, when uncovered and "observed," to be 131. The largest error of "prediction" is 24 points. This is the famous and often cited case of Gladys, IQ 92, and Helen, IQ 116, in the twins study of Newman, Freeman, and Holzinger,[5] one of the four studies in the Jensen compilation mentioned above.

The Las Vegas method of analysis of variance in figure 2(a) creates a twin pair with six cards: Draw four cards from the S–N50–V40 deck and add their integers with due regard to sign; the sum is disturbance from the population norm of 100 due to genetics that is common to both twins of the pair; draw one more card for the environment of one twin and add this to obtain that twin's IQ. Draw one more and do the same for the other twin. Genetic cards have four times more influence than environmental cards on each individual's IQ. An example: The highest "predicted" IQ of figure 2(a) had a sum of 31 for genetics plus three for environment for an IQ of 134, and the other "observed" twin had zero for environment for a total of 131. (A perfectionist shuffles after each card draw, although this is not really necessary; just put drawn cards back at random between twins.) For 60 percent geneticity, use seven cards; three for common genetics and two two's for environments.

On what basis are the obvious results of figure 1 rejected? And they are rejected—believe me! Let me quote from a recent letter signed by a past president of the American Psychological Association in response to an inquiry a friend made about my reasoning:

"When Dr. Shockley says that heredity is more than twice as important as environment in determining the IQ, he doesn't know what he is talking about and doesn't understand the problem. Both variables are completely important. Any other statement is nonsense."

I have failed to detect any impressive capacity for analytic thinking behind such dogmatic assertions. I shall give two examples of the feeble thinking that accompanies the rejection of the "more-than-twice-as-much" conclusion drawn from figure 1.

Here is a typical statement concerning my first example: "Identical twins are not absolutely identical. After all, nature must make occasional errors in perfect duplication of genes. The analysis of figure 1 does not allow for such differences. Consequently, the deductions

---

[4] Arthur R. Jensen, "IQ's of Identical Twins Reared Apart," *Behavioral Genetics*, No. 2, 1970, pp. 133–46.

[5] H. H. Newman, F. N. Freeman, and K. J. Holzinger, *Twins: A Study of Heredity and Environment*. Chicago: University of Chicago Press, 1937.

may be in error. Until you know how much error, the conclusion that geneticity is 82 percent may be way off. It might really be less than 50 percent if the genetic accidents were large enough."

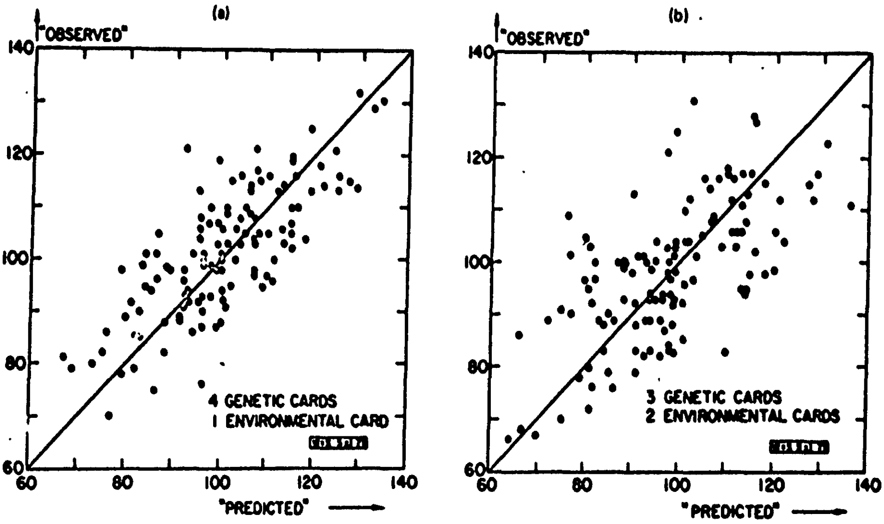


FIGURE 2.—The challenge continued: By "creating" artificial individuals with randomly generated deviations from the population norm of 100 IQ, scatter diagrams like figure 1 are made, (a) Four parts genetics and one part environment is seen to match the real data of figure 1. (b) Three parts genetics to two of environment gives less IQ predictability than is actually found.

I have heard this ridiculous argument seriously proposed by presumably competent biologists. I introduced it as a sort of IQ test for a group of able science writers at a seminar on the Las Vegas method; none of them got it. On another occasion I tried it on a group of Stanford biology majors; it was shot down by a freshman while an upperclassman remained baffled until after the answer was explained twice. Here is the answer:

If geneticity were really 80 percent but accidental gene duplication errors caused many of the twins to differ by, say, 10 IQ points, then this difference would not be allowed for in plotting figure 1. Consequently, the error of prediction would be increased due to the unknown genetic differences. We would attribute these additional errors to environment. In other words, the effect would be to make us wrongly overestimate the effects of environment and underestimate geneticity. Thus if the neglected effects are really present, correcting for them could not lead to a lower correct value like 50 percent but only to a higher value than 80 percent.

Another standard argument for rejecting genetic dominance of IQ asserts that IQ is really controlled by environment; IQs of separated identical twins are nearly equal because adoption agencies succeed in placing the two twins of a pair in essentially identical environments. This "equivalent-environment" argument does not stand up against the facts. The best data is that of the late Sir Cyril Burt, whose 1966 paper[6] supplied 53 of the pairs of twins in figure 1. I

[6] Cyril Burt, "The Genetic Determination of Differences in Intelligence: A Study of Monozygotic Twins Reared Together and Apart," *British Journal of Psychology*, 1966, pp. 137–53.

had obtained these values from Sir Cyril to construct possibly the first scatter diagram plot like figure 1, thinking that the raw data would be a more eloquent witness to the realities of human intelligence than the usual tabulations of correlation coefficients. In response to my subsequent inquiries, Sir Cyril reviewed his reasons for refuting the equivalent-environment explanation. I select for my example of his comments the one on the previously mentioned 132–131 pair of figure 1. About these twins he wrote:

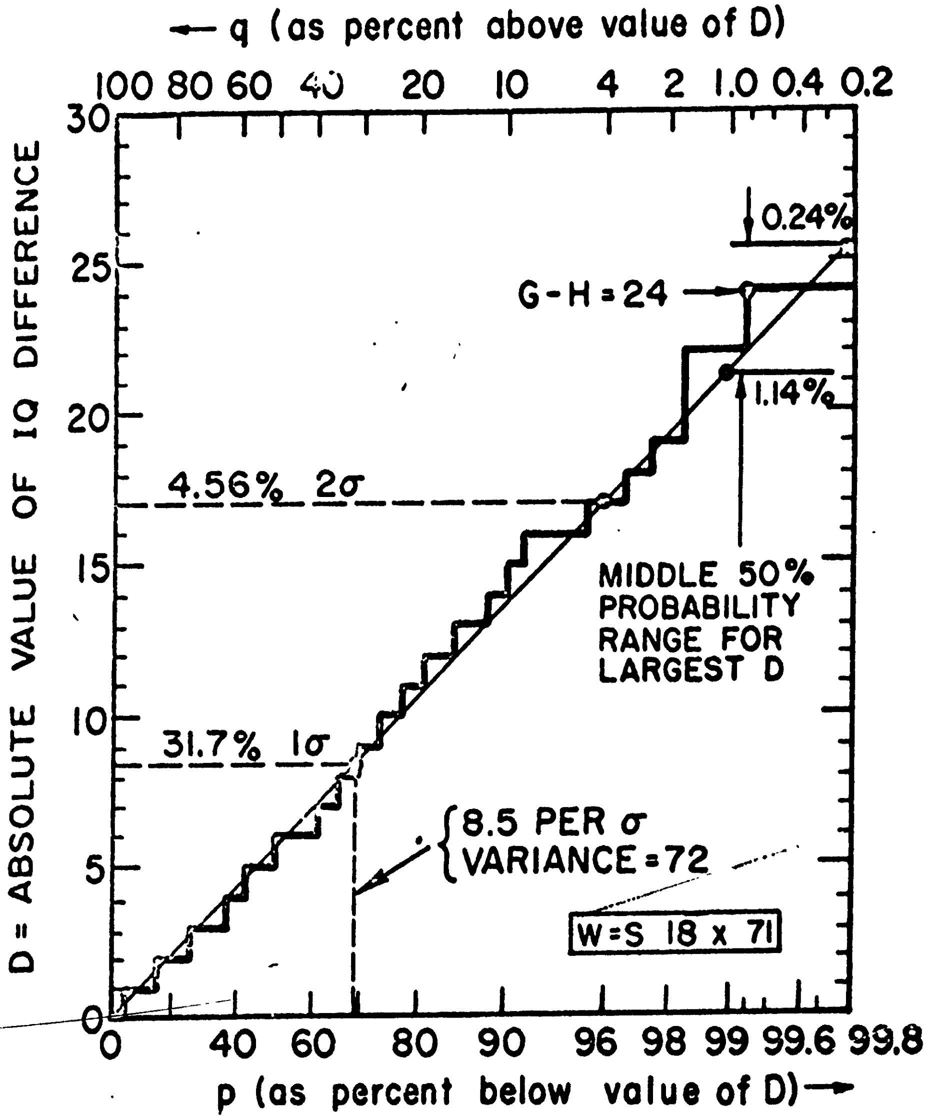


FIGURE 3.—Nongenetic influence are seen to cause IQ differences to be accurately normally distributed. [Dear reader: If you are responding to my challenge, don't spoil my detective story by reading the answer in the text that explains this figure now!]

"They were children of an Oxford don [Burt rates this as occupational class '1,' the highest of the six he lists for home environments] who died a few months before their birth. Unable with her slender

means to bring up two boys as she would desire, [the widow] secretly arranged for one to be 'boarded out': He was sent to a farmer in Wales (occupation class '4') and eventually became a successful farmer himself (Miss Conway gives his IQ in 1958 as 137; our final assessment was 132). The one who remained with his mother eventually obtained a first class degree (IQ 136 in 1958, 131 in 1956)." [7]

This quotation illustrates two general conclusions of Burt's study: There is no significant correlation—indeed, the correlation coefficient is slightly negative—between the environments of Burt's separated twins. It also illustrates the typical range of test errors that may occur—on the order of five points. In the carefully controlled tests used in the four twin studies compiled by Jensen, test error is estimated to be normally distributed with a standard deviation of about 3.5 points so that it contributes about 5 percent, or 10 units, to the population variance of 200.

If the differences in environments between pairs of twins are compared with their differences in IQ for Burt's compilation, then it turns out—as makes sense—that better occupational class of home does tend to raise IQ—but this tendency is not a certainty nor are the IQ increases very decisive: Of the 35 cases in which co-twins differed in both IQ and occupational class, 23 were concordant—higher class with higher IQ—and 11 were discordant—lower IQ in the higher class home. The result is significant at the 0.02 level. Each upward step of one social class raises IQ on the average about one IQ point.

But what about Gladys and Helen, with their 24-point difference? The difference is often cited to show that environmental effects among Caucasions are so much larger than differences between racial averages that obviously environment can easily account for the generally accepted deficit of about 15 points of our Nation's black minority. The Gladys-Helen case warrants close scrutiny.

The Gladys-Helen 24-point difference is the exception needed to prove the 80-percent geneticity rule: It would be improbable if there were not one such case with a difference of about 24 IQ points in a sample of 122 pairs of twins. The reasoning is outlined on figure 3. In brief, the method of plotting shows that the differences (D) in IQ between twins is as accurate a realization of a normal distribution as one could expect from 122 cases. Therefore, although we may not be able to identify what the exact causes are that push the IQ of one twin away from the IQ of his cotwin, there are apparently enough independent, additive causes to give a good normal distribution. If it is a normal distribution, then straightforward methods can be used to determine the range of IQs in which the highest of the 122 differences has a 50-percent chance of falling—the probability being 25 percent that the largest falls above and 25 percent that it falls below this range. Gladys-Helen does fall in the proper range, as shown on figure 3. There is only one chance in 100 that the largest value would have been smaller than 17 points.

One more logical consequence of figure 3 is that one standard deviation of the environmental variable that influences IQ is worth five IQ points. Even though we cannot define what this variable may be—

[7] Sir Cyril Burt, personal correspondence with the author.

undoubtedly it is some complex combination of many components—it must account for some 25 units of variance for each twin to give the standard deviation of 8.5 in figure 3 in combination with test error variance. Burt's occupational class variable only accounts for about one-fourth of this unknown environment composite.

Applied to Gladys-Helen, this five-point environmental variable accounts for a large fraction of the 24-point difference: Gladys had a sickly childhood and never finished third grade. Helen graduated from college. This large environmental difference, appraised using Census Bureau tables, corresponds to quite possibly three or four standard deviations of the distribution of educational environments—the 80 percent geneticity model can thus account for a substantial fraction of the 24 point difference. As Herrnstein's recent widely noted article in the Atlantic emphasizes,[8] if such large environmental differences were eliminated by social progress, then the relative importance of genetic differences would increase.

One final significant point about figure 3 and the accurate 82 percent geneticity value that can be deduced from it in conjunction with figure 1: If the true value for geneticity were as small as 72 percent, then standard statistical theorems lead to the result that there is less than one chance in 2,000 that a value as small as the 8.5 for the standard deviation of figure 3 would have occurred by chance.[9] This is a typical level of significance statement. It says that the hypothesis that geneticity is 72 percent or less can be rejected at a significance level of 0.0005 so far as the null hypothesis that 8.5 of figure 3 resulted by chance is concerned.

THE NON-GENETIC 20 PERCENT

My emphasis on the dominance of genes in controlling IQ has led to the misunderstanding that I "treat IQ as a fixed characteristic, like eye color, susceptible of exact measurement"—to quote from an editor's reaction to one of my manuscripts. A distinguished psychologist, after seeing a diagram showing environmental effects based on the 80 percent geneticity presented above, wrote to me: "Your figure implies that no matter how bad the environmental restriction becomes it will have no effect whatsoever on the phenotype indicated by the IQ test score. This would mean that if William Shockley had been raised in a clothes closet from the time he was old enough to learn language, he would still have been able to win the Nobel Prize."

The fact is that, as for the Gladys-Helen case, small though the 12 to 15 percent of the variance attributable to environment may be, it can have large effects upon IQ and other behavioral traits. In fact, some of my own educational experiments have been aimed at raising IQ or

---

[8] Richard Herrnstein, "IQ," *Atlantic Monthly*, September 1971, pp. 43–64.

[9] William Shockley, "On the Significance Level for Genetic Dominance of IQ and on the 24-Point Difference Between Twins Gladys and Helen," paper presented at October 27 meeting, 1971, of the National Academy of Sciences, Washington, D.C.

motovational or attitudinal factors. Figure 4 illustrates one surprisingly successful result. For a number of years my freshman seminar at Stanford was chosen by almost twice as many students as I could take in two sections. I rated them in groups having closely matched weighted averages of S.A.T. scores and from each matched group rejected about half by using random numbers. The experimental group was found to outperform the controls by about 0.6 of a standard deviation of grade point average for the four academic quarters subsequent to the two spent in the seminar.

A recent widely publicized example of exceptional environmental success in reducing mental retardation may fit into the 80 percent geneticity pattern. Prof. Rick Heber has given an intensive educational enrichment program to slum children whose mothers have IQs below 75. At 3½ years of age, the undersecretary of Health, Education, and Welfare has recently reported, these experimental children are averaging 33 IQ points above comparable controls.[10] These findings are not incompatible with 80 percent geneticity. In fact, they may be almost predictable. The undisturbed home environments were probably in the lowest 1 or 2 percent of all home environments for intellectual stimulation. On the other hand, Heber's intensive program is probably in the top fraction of 1 percent for developing performance on IQ tests. This is equivalent to an improvement of perhaps six standard deviations of the distribution of environments, so that 33 points would correspond to about five points per standard deviation—a value quite compatible with 80 percent geneticity.

The economics of such remedial programs suggest mournful numbers. The initial cost was of the order $10,000 per child year. Whether the effects will be lasting or in the end adverse because of untimely experiences—such is the case for laboratory experiments with primates—are important and researchable questions.[11] I discuss below the moral obligation to do quantitative thinking on human problems.

STANDARD IQ CLICHES

I have gone at length and with dramatized examples into the basis for my own conviction about genetic dominance of IQ because I believe that this is the cornerstone for all logical structures about human quality problems. I anticipate that many criticisms will be leveled at my reasoning. Some of these I shall respond to in detail below. Here I shall deal perfunctorily with some that space does not permit me to treat in depth:

---

[10] John G. Veneman, "Partial Text of Remarks by Undersecretary John G. Veneman, Department of Health, Education, and Welfare, before the Pacific Forum on Mental Retardation, Honolulu, Hawaii, September 29, 1971." Released by Office of Public Affairs, HEW, Washington, D.C.

[11] Arthur R. Jensen, "Reducing the Heredity-Environment Uncertainty," in *Environment, Heredity, and Intelligence*, Reprint Series No. 2, *Harvard Educational Review*, Cambridge, Mass., p. 234. (Jensen's discussion is based on H. F. Harlow, "The Development of Learning in the Rhesus Monkey," *American Scientist*, 1959, pp. 459–79.)

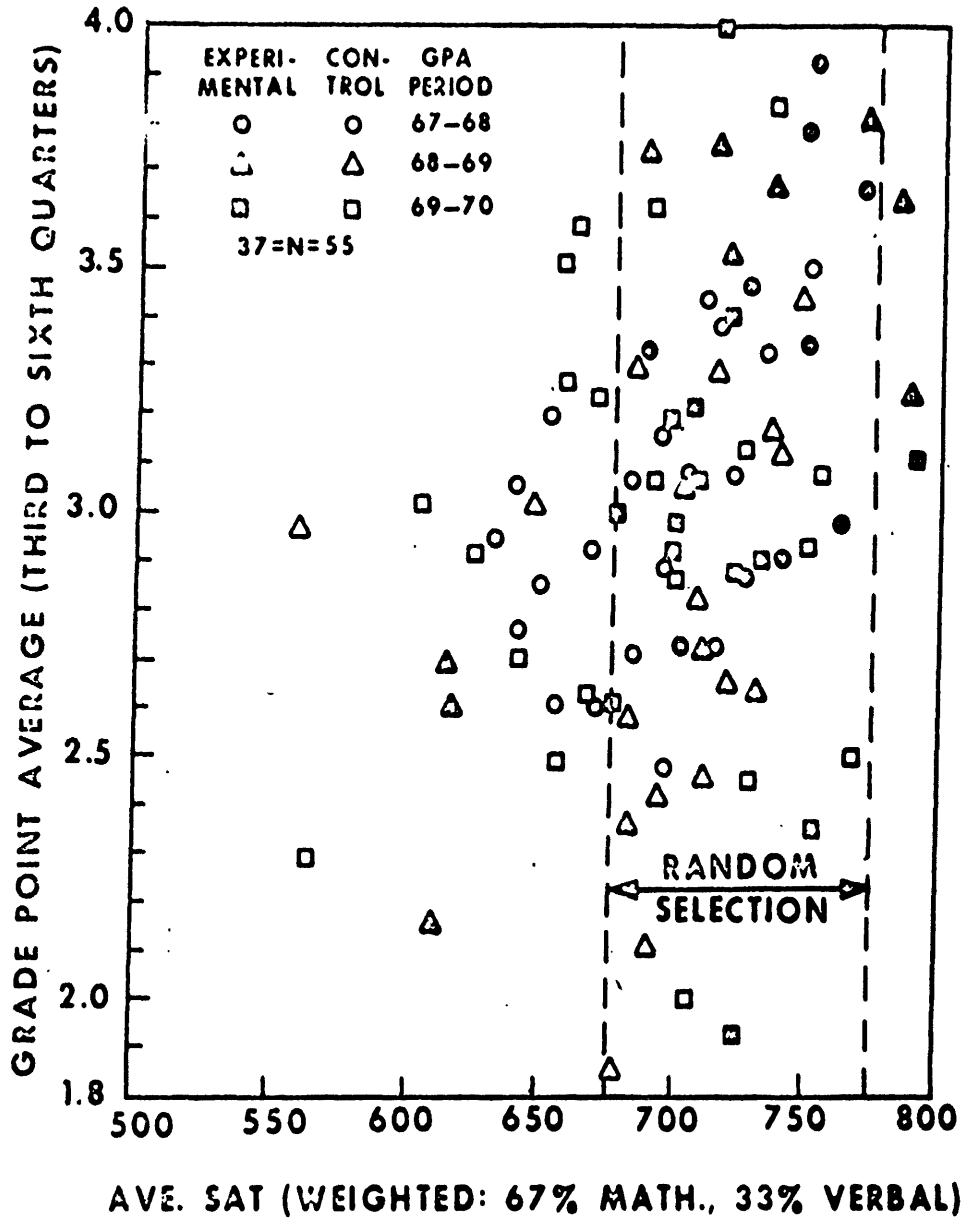


FIGURE 4.—Results of a controlled experiment on randomly selected applicants for a Stanford University freshman seminar on mental tools for scientific thinking. In the four quarters subsequent to the two in which the seminar was taken, the "experimental" students outperformed the controls at a significance level of better than 0.05.

"IQ has no relevance to successful living." My best answer to this is an analysis of the 'Genetic Studies of Genius,' the great work of Lewis M. Terman and his colleagues. The gifted children did outperform the population average across the board on all sorts of generally accepted and valued human quality measures.[12]

[12] Lewis M. Terman and Melita H. Oden, *The Gifted Group at Mid-Life: Thirty-five Years' Follow-up of the Superior Child*, vol. 5, Stanford, Calif.: Stanford University Press, 1959.

"Until you can meaningfully define exactly what you mean by intelligence and relate it to what IQ measures, your studies are not scientific." My answer is that IQ as used by Terman and others is meaningfully correlated with values that are generally accepted. I also turn the question: Until you can tell me what is gravity, why should I worry about falling?

"IQ tests are so culturally influenced that they cannot possibly tell anything about genetic potential and especially about racial differences. For example, monkeys could outperform humans on tests involving tree climbing." One answer that almost always reveals the unsearch dogmatism of the questioner is this: What is the best attempt that you know of to design a culture-fair test and what was wrong with it? I do not recall ever getting an answer. I shall discuss several research proposals on racial differences below.

"You have discussed geneticity; but what does this have to do with dysgenics—after all, dumb parents have bright kids and vice versa?" My answer: See any good psychology text on correlation of adopted children's IQs with natural and with foster parents. In fact, these comparisons are the independent way to arrive at the 80 percent geneticity figure without using identical-twin data. Let me express the conclusion by quoting again from the Sir Cyril Burt letter mentioned above: "But the strongest case for mental inheritance is provided by a comparison of data for all types of relatives."

The list is long. It may have no end. The "Apple of God's Eye Obsession" may drive true believers tirelessly. For other answers I refer my readers to the references, particularly Jensen,[13] Eysenck,[14] Herrnstein,[15] and my own writings with their reference lists.[16]

FORMS OF DYSGENIC THREAT

My concerns are based on my evaluation that in the intellectual community of the nation the emphasis on environmental aspects of human quality is so great that it excludes proper consideration of hereditary genetic factors. I appraise this unbalance as deplorable and dangerous. During the last half decade my studies have increased my conviction that concentration upon the environmental factors cannot solve the important problems of man's future and that adequate solutions to poverty, crime, illiteracy, and national security problems demand facing hereditary problems. I believe that to avoid very real dangers to worldwide human welfare, civilization, including particularly that of the United States, must face in a broader sense than it does now the problems raised in 1966 by James Shannon, then director of the National Institutes of Health, in congressional testimony: "The effect—if I may put it bluntly, Mr. Chairman—is that we are weakening our genetic inheritance."[17] Dr. Shannon emphasized biochemical physiological

---

[13] Arthur R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?", *Harvard Educational Review*, Winter, 1969, pp. 1–123.

[14] H. J. Eysenck, *The IQ Argument*, Freeport, N.Y.: The Library Press, 1971.

[15] Herrnstein, *op. cit.*

[16] William Shockley, "Models, Mathematics, and the Moral Obligation To Diagnose the Origin of Negro IQ Deficits," *Review of Educational Research*, October, 1971, pp. 369–77.

[17] James A. Shannon, testimony before House Subcommittee on Appropriations, March 2, 1966. See *Washington Star*, March 25, 1966, p. 1.

traits. What my intellectual conscience impels me to demand is that we look objectively also on man's behavioral traits. This, my investigations lead me to conclude, is not being done adequately. I conjecture that this lack of needed effort is caused less by the great difficulties involved than by the unsearch dogmatism that produces thought-blockers.

With the advent of nuclear weapons, man has in effect reached the point of no return in the necessity to continue his intellectual evolution. Unless his collective mental ability can enable him reliably to predict consequences of his actions, it is possible that he may provoke his own extinction, or at least drastically modify the gene pool of humanity—and perhaps for the better.

Let me illustrate by a specific speculation upon the evolutionary aspects of possible gene pool modifications: Sweden and Switzerland both have extensive shelter facilities that would save substantial fractions of their populations from death from worldwide fallout in the event of an unlimited nuclear war involving "dirty weapons" that might destroy the preponderance of the human life on earth. A much less substantial fraction of our population would survive. This preferential survival of the most foresighted components of the human race is a form of "self-renewal" for human evolution that my intellectual conscience does not allow me to face complacently. I feel an obligation to try to increase the probability that man's destiny will be shaped by the application of intelligence to determine realistic goals for human progress rather than by forces man has let get out of control. These speculations about man's future evolution accent my fears that contemporary U.S. population trends are such that we are disproportionately multiplying the least foresighted elements of our population.

A nuclear holocaust as a consequence of advancing weapons technology combined with a dysgenic decline in national foresight may present the most dramatic dysgenic threat. But increased welfare tax burdens and crime rates and lower productivity may act sooner to draw attention to the basic issues. I estimate that our nobly intended welfare programs may be encouraging the births of 100 babies per day who can be reliably predicted to face lives of frustration because of low genetic IQ potential. It is this estimate—I find no one in Government who will check it—as much as any one thing, that underlines the urgent need for evaluation. I propose as a program for continued progress: Let's ask the questions, do the necessary research, get the answers, discuss them widely. Then either worries will evaporate or plans for action will develop.

RACEOLOGY

A common objection to studies of racial genetics is that the concept of race is meaningless. This objection is refuted by research on blood type frequencies, most recently that of T. E. Reed of Toronto, who has determined with a precision of 1 percent that the Oakland, Calif., Negro population is 22 percent Caucasian in ancestry.[18] I have refined

---

[18] T. E. Reed, "Caucasian Genes in American Negroes," *Science*, August 22, 1969, pp. 762–68.

Reed's studies to estimate that the spread of the Caucasian ancestry in Oakland probably varies from a few percent to well over 50 percent [19] and have combined Reed's findings with Army pre-induction test data in figure 5 to estimate that, for low IQ Negro populations, each 1 percent of Caucasian ancestry raises average IQ by one point.[20] I have suggested ways of controlling for the environmental differences to test the reliability of this estimate. An interesting question is the level at which diminishing returns set in; for example, at 40 percent Caucasian ancestry, would average IQ be 110?

The possible relationship of blood type determination of racial mixes of populations and IQ may offer a unique opportunity to evaluate the reality of the dysgenic threat. To fail to use a potentially effective means of diagnosis for fear of being called a racist is irresponsible. It may also be a great injustice to black Americans themselves. If those Negroes with the fewest Caucasian genes are in fact the most prolific and also the least intelligent, then genetic enslavement will be the destiny of their next generation.[21] The consequences may be extremes of racism and agony for both blacks and whites.

The word "raceology" has been proposed for studies like mine. They are not racist. They are motivated by concern for the feelings of all involved—not by fear and hate. My research focuses principally upon white-Negro comparisons for two reasons: (1) Our national racial problems primarily involve the Negro minority and (2) Negroes are the only racial group for which extensive published statistics are available. Therefore, my personal research on questions related to Negroes has far greater immediate promise of contributing to sound diagnosis of our human quality problems than, for example, would attempts to study hereditary factors for Appalachian whites, for whom I have found that statistical data are practically unobtainable. Although I emphasize the Negro area for these reasons, I continue to urge broad inquiry into hereditary aspects of human behavior for all racial groups.

As an example of raceology, I present in figure 6 some new research results on Negro superiority that compare Negro and white visual acuity, based on Army tests. The points specify fractions of Negroes

---

[19] William Shockley, "Hardy-Weinberg Law Generalized To Estimate Hybrid Variance for Negro Populations and Reduce Racial Aspects of the Environment-Heredity Uncertainty," Proceedings of the National Academy of Sciences, vol. 68, 1971, p. 1390A.

[20] William Shockley, "New Methodology To Reduce the Environment-Heredity Uncertainty About Dysgenics," Proceedings of the National Academy of Sciences, vol. 67, 1970, pp. 10A–11A (abstract):

[21] That dysgenics is more threatening for Negroes follows from D. P. Moynihan, "Employment, Income, and the Ordeal of the Negro Family," in *The Negro American*, T. Parsons and K. B. Clarks (eds.), and B. T. Osborne, "Population Pollution," *Journal of Psychology*, 1970, pp. 187–91. Moynihan reports that "in 1960 nonwhite women (married once, husband present) age 35 to 45 had 4.7 children as against 3.8 for white women in the same situation" (p. 148). For women in the same age bracket, married at age 22 or over to professionals or technical workers with one or more years in colleges, the numbers are 1.9 children for Negroes and 2.4 for whites. Osborne reviewed the standard treatments that reject all evidence for dysgenic trends. He presented new findings and came to the conclusion that prior studies were based on populations too narrowly selected and that dysgenic trends cannot be soundly rejected.

and whites having various levels of visual acuity. From 20/20 to less than 20/200, the points fall accurately along a line. The interpretation of this analysis is that whites and Negroes are distributed in their visual acuity according to the same basic underlying normal distribution but that the distribution for Negro visual acuity is offset upwards by approximately 0.6 of a standard deviation—a value that if it applied for mental performance would be equivalent to about nine IQ points.

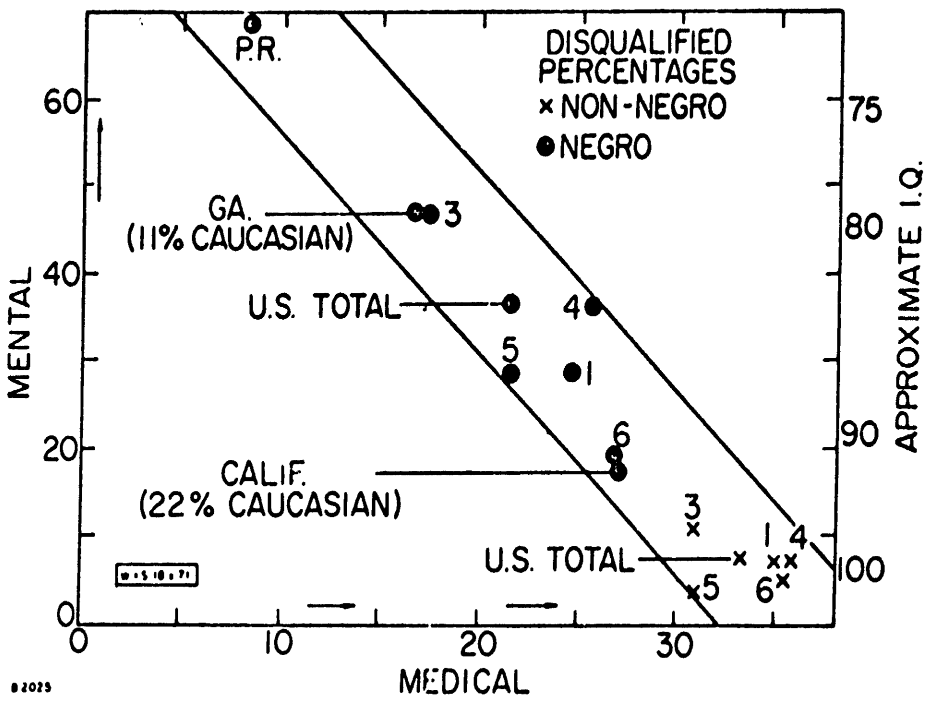


FIGURE 5.—A plot of data from *Supplement to Health of the Army,* June 1969 (Bernard D. Karpinos, Medical Statistics Agency, Office of the Surgeon General, Department of the Army). The data apply to preinduction examinations of draftees in 1968. The numbers identify five recruiting districts; P.R. is Puerto Rico. Caucasian percentages are from T. E. Reed's values for Oakland, Calif., and two counties in Georgia. Mental and medical rejection rate coordinates include those rejected on both grounds. The approximate IQ scale is obtained by assuming a normal distribution with a standard deviation of 15 and 100 IQ for non-Negro U.S. total. (For citations, see footnotes 18 and 19.)

Medical studies support the conclusion that the differences between the Negro and the white distributions of visual acuity are due to differences in gene pools rather than environmental effects. This shoots down the theory of some social scientists that many white children ruin their eyes by excessive reading and that this is why white visual acuity is worse than black. The opinion of ophthalmologists is that myopia, the chief cause of poor visual acuity, does not arise from excessive use of eyes for close work such as reading. Large-scale studies extending over periods of years have prevented children from focusing at short distances by mild doses of atropine that are known not to affect normal eyes. The subjects were expected to develop myopia in a certain percentage of cases on the genetic basis that their families had high

incidence of myopia. No reduction of myopia was found. The fact that gene pool effects are involved is further supported by the dominance of myopia over hypermetropia, or farsightedness, in studies of family patterns of poor vision.[22]

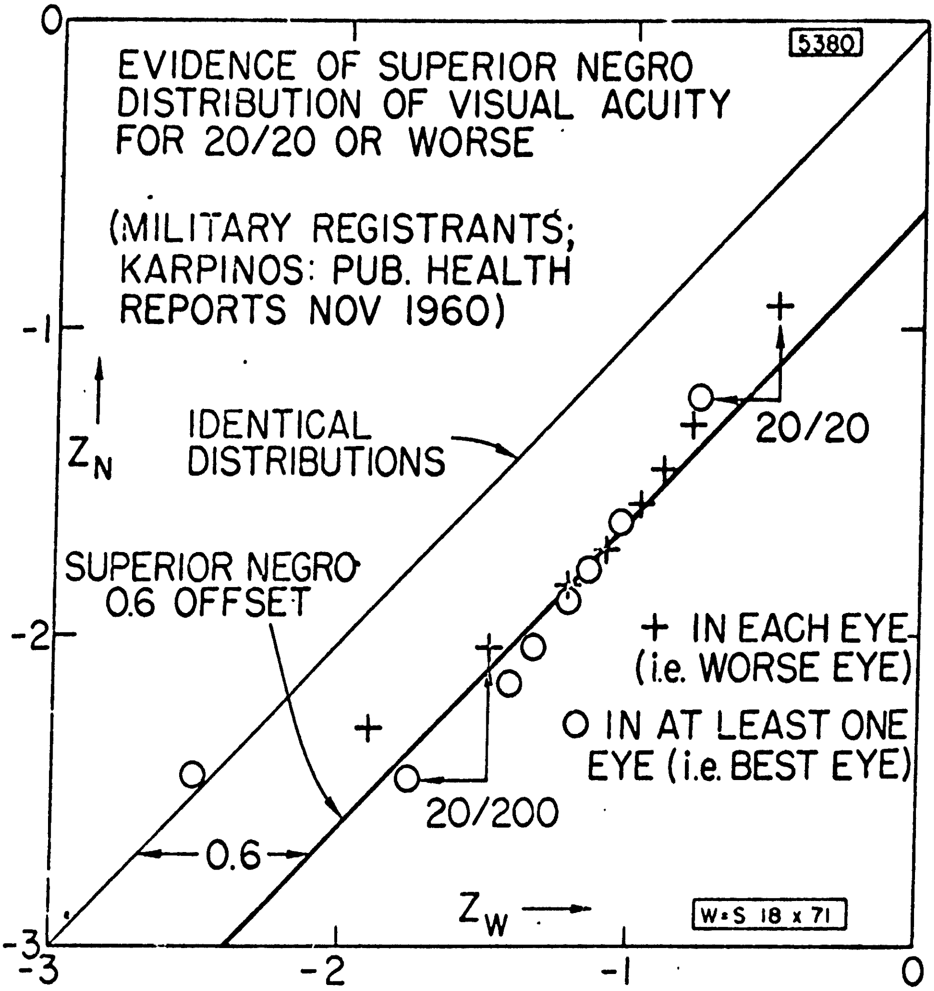


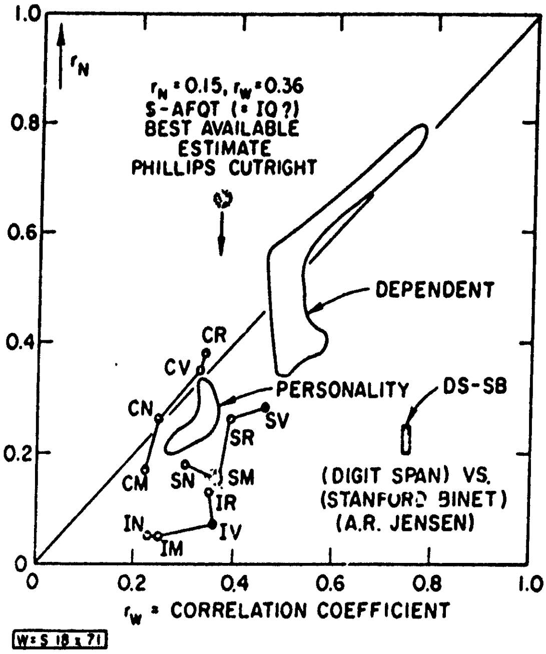
Figure 6.—The Z_w values give normal distribution arguments that correspond to the percentage of white military registrants who fail to meet the prescribed visual acuity. Z_N values are plotted in the same way for Negroes. The unmarked visual acuities are in sequence 20/30, 20/40, 20/50, 20/70, 20/100. The extreme points that fall out of the pattern are 20/400. If the points fell perfectly on the line, it would imply identical normal distributions for both races except for an offset of 0.6 standard deviations.

Correlation coefficients between behavioral traits were found to be smaller for Negroes than for whites[23] using data from tables in the Coleman Report.[24] Figure 7 presents these data so as to facilitate in-

[22] John B. deC. M. Saunders, personal communication to the author, based on his review of the problem at the University of California Medical School in San Francisco.

[23] William Shockley, "'Cooperative Correlation' Hypothesis for Racial Differences in Earning Power," Proceedings of the National Academy of Sciences, Vol. 66, 1970, p. 245 (abstract).

[24] James S. Coleman et al., *Equality of Educational Opportunity.* Washington D.C.: Department of Health, Education, and Welfare, 1966.



CORRELATION COEFFICIENTS
"COLEMAN" REPORT GRADE XII

DEPENDENT=ACHIEVEMENT
V=VERBAL
N=NONVERBAL
R=READING COMPREHENSION
M=MATHEMATICS ACHIEVEMENT

"PERSONALITY"=STUDENT
S=SELF CONCEPT
I=INTEREST IN SCHOOL
C=CONTROL OF ENVIRONMENT

	S	*I*	*C*	*V*	*NV*	*R*	*M*
W Av.	0.02	-0.03	0.22	67	37	76	57
W. S.D.	0.73	0.69	0.57	14	7	17	17
N Av.	0 06	0.11	-0.13	49	28	58	38
N S.D.	0 71	0.55	0.76	14	9	17	18

FIGURE 7.—Comparison of correlation coefficients r_W for whites and r_N for Negroes for correlations between achievement variables and personality variables. The lower "cooperative correlation" is consistent with the Cutright estimate of lower effect of IQ on earnings for Negroes than for whites. (Phillips Cutright, personal communication to W. Shockley, September 22, 1969.)

terracial comparisons of the correlation coefficients between "student" variables and "dependent" or achievement variables. As figure 7 shows, except for the remarkable "control-of-environment" variable, the correlation coefficients between student variables and achievement variables are much lower for Negroes than for whites. The mean values of such student "personality" variables as self-concept and interest in school are seen to be no lower for Negroes than for whites—in fact, they are slightly higher for Negroes. What is surprising, however, is the difference in the pattern of correlations between the personality variables and the achievement variables. Comparisons between Orientals and whites do not show the striking differences in values of correlation coefficients. Explanations of the lower correlation between IQ and earnings for Negroes than for whites usually lean heavily on the fact that blacks in our society are subject to racial discrimination. I have used my findings to offer an explanation of the lower correlation not involving discrimination.[25] The differences shown here are consistent with differences in level I (rote memory) and level II (conceptual) learning reported by Jensen.[26] The chief purpose in introducing figure 7 here is to illustrate the existence of research possibilities on racial differences that may exist but are unexplored because of the prevailing unsearch dogmatism.

---

[25] William Shockley, "Dr. Shockley's Theory," *The New York Times*, Nov. 8, 1969, p. 32C.

[26] Jensen, *op. cit.* (fn. 13), p. 122.

Where data have been available, I have tried to compare other racial groups.[27] My findings do not support a theory of white Aryan supremacy: I have found and published the observation that American Orientals are about 10 times more successful than the national average on a per-capita basis in achieving the distinction of election to the National Academy of Sciences. They are also about 10 times more successful in avoiding citations in the annual FBI uniform crime reports.[27] My statistics also show that Jewish Nobel Prize winners in science occur about 10 times more often than expected on the basis of the population as a whole..

QUANTIFIABLE HUMANISM?

One form objections take to my demands that quantitative scientific thinking be applied to human quality problems was eloquently expressed in a listing of and comment on environmental variables in a letter by a black Ph. D. in education as part of his criticism of a paper of mine:

"* * * devastation * * * has been wreaked * * * through the evils of slavery, * * * intimidation, lynching, virulent job discrimination, segregation, * * * How can the debilitating effects of such a legacy be couched in quantifiable terms?"

I believe we must answer that we do not, nor shall we soon, know how to quantify such environmental factors. But the future of our Nation's black minority does depend upon sound diagnosis. Wishful thinking and good intentions are not enough. Quantified facts do describe the agonizing disadvantages of Afro-Americans. Note this recent Associated Press dispatch:

"The NAACPs labor director, Herbert Hill, told the annual convention: 'The rates of unemployment among black youth have now reached disaster levels. And if they continue * * * virtually an entire generation of ghetto youth will never enter the labor force. Their only future will be a marginal, alienated existence, separate and unusual within American society * * *' "[28]

Mr. Hill's concern over black unhappiness is supported by a Gallup poll of 1,517 adults. "Very happy" was the response of 46 percent of whites but of only 20 percent nonwhites; "not happy" percentages were 5 percent and 12 percent.[29]

What do these quantitative findings mean? My "offset analysis"[30] of these percentages shows that the nonwhite happiness distribution is offset downwards, compared with whites, by about half a standard deviation for adults. What will it be for the next generation of black Americans whose employment disaster Hill reports? Will diagnosis reveal that racial dysgenics is a cause? Diagnosis of questions like those related to Negro unhappiness is what I believe will be the best insur-

---

[27] William Shockley, "A 'Try Simplest Cases' Approach to the Heredity-Poverty-Crime Problem," Proceedings of the National Academy of Sciences, Vol. 57, June 1967, pp. 1767–74.

[28] Associated Press, *San Francisco Chronicle*, July 9, 1971, p. 7. Cited by Shockley in *op. cit.* (fn. 16).

[29] George H. Gallup, "Gallup's Poll: The Happiest People," *San Francisco Chronicle*, January 14, 1971, pp. 1–2.

[30] William Shockley, "Offset Analysis Description of Racial Differences," Proceedings of the National Academy of Sciences, Vol. 64, 1969, p. 1432 (abstract).

ance for our black minority's future and what I urge our Nation's citizens, including the professional educators who read this journal, to demand.

THE MORAL OBLIGATION TO THINK

1. Hitler and Speer. A familiar basis for rejecting my demands that research on dysgenics be undertaken is the assertion that any resulting knowledge would be worthless because all conceivable remedial action would involve intolerable eugenic measures.

Eugenics is a shunned word because it was a feature of Hitlerism. But the lesson of Nazi history is not that eugenics is intolerable. Since 1935 Denmark has carried out programs with clearly positive eugenic implications. (Although a cause-and-effect relationship is uncertain, it is noteworthy that Denmark's per-capita homicide rate has dropped since World War II and is less than 2 percent of the rising rate for Washington, D.C., which was 20 percent higher in 1971 than in 1970.) The real lesson of Nazi history was anticipated 140 years before Hitler, when the Bill of Rights incorporated into our Constitution the First amendment guaranteeing freedom of speech and of the press. Only the most anti-Teutonic racist can believe the German people to be such an evil breed that they would have tolerated the concentration camps and gas chambers if a working First Amendment had permitted exposure and discussion of Hitler's final solution—the extermination of the Jews.

I suggest that there is a significant parallel between the attitude of German intellectuals in Hitler's day and our intellectuals' unwillingness to face the dysgenic threat. Albert Speer, Hitler's minister of armaments and war production, wrote in his memoirs:

"But in the final analysis I myself determined the degree of my isolation [from Hitler's "final solution" of the Jewish problem], the extremity of my evasions, and the extent of my ignorance. . . *Whether I knew or did not know, or how much or how little I knew, is totally unimportant when I consider what horrors I ought to have known about and what conclusions would have been the natural ones to draw from the little I did know.* Those who ask me are fundamentally expecting me to offer justifications. But I have none. No apologies are possible." [31] [Emphasis added.]

I call this retrospection the "Speer syndrome." It is what I warned Kappan readers who failed my challenge that they might experience in future decades if—to paraphrase Speer—they are failing to draw the natural conclusions from the little—or much—they do know.

2. A voluntary sterilization bonus plan. The First Amendment makes it safe for us in the U.S. to try to find humane eugenic measures. As a step in such search, I propose as a thinking exercise a voluntary sterilization bonus.

Bonuses would be offered for sterilization. Payers of income tax would get nothing. Bonuses for all others, regardless of sex, race, or welfare status, would depend on best scientific estimates of hereditary factors in disadvantages such as diabetes, epilepsy, heroin addiction, arthritis. At a bonus rate of $1,000 for each point below 100 IQ,

---

[31] Albert Speer, *Inside the Third Reich: Memoirs of Albert Speer.* New York: Macmillan, 1970, p. 113.

$30,000 put in trust for a 70 IQ moron potentially capable of producing 20 children might return $250,000 to taxpayers in reduced costs of mental retardation care. Ten percent of the bonus in spot cash might put our national talent for entrepreneurship into action.

In Honolulu on September 29, 1971, John G. Veneman, Under Secretary of Health, Education, and Welfare, rejected this thinking exercise, saying:

"And the more I thought about [the voluntary sterilization bonus plan], the less I liked that idea. All my instincts told me that the way to attack mental retardation is at its roots—not through its victims. For many years I was a fruit grower in California. And I've learned that you begin with good rich soil—not with the fruit. * * *" [32]

He did not mention seed quality. This substitution of instinct for scientific analysis and emphasis on environmental soil to the exclusion of genetic seed quality reminded me of Lysenko in Russia. With Stalin's backing, he insisted that his Soviet biologists had discovered how to transform one species into another—wheat into rye, pines into firs, et cetera. Lysenkoism was a disaster in Russian agriculture.

One obvious area of tabooed research comparable in emotional hazard to conventional genetics in Lysenko's Russia, concerns racial differences in brain anatomy. The most significant recent publication that I can find reports "unexpected variations in fine structures of the brain in Melanesians, including size and shape of septal nuclei, * * * and the frontal lobes." [33] Where has this research on racial frontal lobe differences, reminiscent of now-rejected research on Negro brain differences, been published? Only in a conference report and an alumni magazine.

Another shocking speculation about dysgenics is provoked by news stories on the "battered child" syndrome. The battered child is becoming more prevalent. Who does the battering? Often it is grown-up battered children.[34] Heritability? Dysgenics?

3. "Apple of God's Eye Obsession." I shall close with a hypothesis about the psychology of the critics of my concerns about dysgenics. I doubt neither the sincerity nor the good intentions of these critics. I diagnose their thought-blockage as caused by a theologico-scientific delusion. I call it the "Apple of God's Eye Obsession"—God meaning, for some, the proper socio-biological order of the universe. True believers hold that God has designed nature's laws so that good intentions suffice to ensure humanity's well-being; the belief satisfies a human need for self-esteem. Any evidence counter to man's claim to be the apple of God's eye strikes a central blow at his self-esteem and thereby provokes retaliation reminiscent of the prompt execution of a Greek messenger bearing tidings of defeat in battle. The parallels become clearer in historical perspective. Galileo and Darwin brought new knowledge that was incompatible with the then-cherished interpretation of humanity's unique place in the universe. Either the new

---

[32] Veneman, op. cit. (fn. 10).

[33] Carleton Gajdusek, "Physiological and Psychological Characteristics of Stone Age Man," Engineering and Science, April 1970, p. 58 (Publication of the California Institute of Technology and the Alumni Association.)

[34] "Parents Who Beat Children," San Francisco Chronicle, August 30, 1971, p. 16.

knowledge had to be rejected or else the Apple of God's Eye Obsession had to be painfully revised.

The thought-blockers and unsearch dogmatism that reject the relevance of genetics to social problems arise, I propose, because the theory that intelligence is largely determined by the genes and that races may differ in distribution of mental capacity offends equalitarian-environmentalism—an important feature of the contemporary form of the Apple of God's Eye Obsession. The preponderance of the world's intellectual community resists the fact that nature can be cruel to the newborn baby. Babies too often get an unfair shake from a badly loaded parental genetic dice cup. At the acme of unfairness are features of racial difference that my own research inescapably leads me to conclude exist: Nature has color-coded groups of individuals so that statistically reliable predictions of their adaptability to intellectually rewarding and effective lives can easily be made and profitably be used by the pragmatic man in the street.

If, as many thinking citizens fear, our welfare programs are unwittingly, but with the noblest of intentions, selectively down-breeding the poor of our slums by encouraging their least foresighted to be most prolific, the consequences will be tragic for both blacks and whites—but proportionately so much worse for our black minority that, as I have said, the consequence may be a form of genetic enslavement that will provoke extremes of racism with agony for all citizens.

My position is that humanity has an obligation to use its intelligence to diagnose and to predict in order to prevent agonies that lack of foresight can all too easily create.

Reprint Series No. 2

Environment, Heredity, and Intelligence

Compiled from the
Harvard
Educational
Review

*Edi. Note: Previously submitted by author with his testimony, this section begins with the Discussion on p. 125.

Discussion: How Much Can We Boost IQ and Scholastic Achievement? #*

The editors have solicited the following discussions in response to Professor Jensen's article, which appeared in the Winter, 1969, issue of the Review. *In his controversial article, Dr. Jensen develops a definition of the concept of intelligence and discusses the relative contribution of genetic and environmental influences in molding IQ. His conclusions run counter to many of the assumptions on which educational programs of the past few years have been based. For example, he argues that IQ is determined much more by genetic than by environmental influences. He argues that the most important environmental factors affecting intelligence occur prenatally and in the first year of life, and are associated mainly with the nourishment of mother and child. He analyzes the failure of many preschool and compensatory programs of the last five years to achieve significant gains in IQ and concludes that such programs are misdirected in their choice of goals and educational practices.*

In the following section six psychologists and a geneticist bring their disciplines and assumptions to bear on his arguments. The Review *welcomes further responses that are pertinent to the issues raised in the article and the discussions. Selected responses will be published in the Letters to the Editor section of subsequent issues.*

* Arthur R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?," Harvard Educational Review, XXXIX (Winter, 1969), pp. 1-123.

Harvard Educational Review Vol. 39 No. 2 Spring 1969

#Previously submitted in testimony, by the author, to the U.S. Senate Select Committee on Equal Educational Opportunity.

77-942 O - 72 - 38

Inadequate Evidence and Illogical Conclusions

JEROME S. KAGAN, *Harvard University*

Professor Kagan is critical of the logic of Dr. Jensen's article and presents evidence that any IQ data collected in the standardized manner may not reflect the actual potential of lower class children. In Kagan's opinion, Jensen's major fallacies are (1) his inappropriate generalization from within-family IQ differences to an argument that separate racial gene pools are necessarily different and (2) his conclusion that IQ differences are genetically determined, although he glosses over evidence of strong environmental influences on tested IQ—even between identical twins. Kagan cites new studies which suggest that part of the perceived intellectual inadequacy of lower class children may derive from a style of mother-child interaction that gives the lower class child less intense exposure to maternal intervention. Finally, Kagan argues, present compensatory education programs have been neither adequately developed nor evaluated. We cannot, therefore, use current evaluations of them to dismiss all possible compensatory programs.

Arthur Jensen's essay on IQ, scholastic achievement, and heredity contains a pair of partially correct empirical generalizations wedded to a logically incorrect conclusion. Professor Jensen notes first that scores on a standard intelligence test are more similar for people with similar genetic constitutions. The more closely related two people are, the more similar their IQ scores, suggesting that there is a genetic contribution to intelligence test performance. The second fact is that black children generally obtain lower IQ scores than whites. Unfortunately, Jensen combines the two facts to draw the logically faulted conclusion that there are genetic determinants behind the lower IQ scores of black children. The error in his logic can be illustrated easily, using stature as an example. There is no doubt that stature

Harvard Educational Review Vol. 39 No. 2 Spring 1969

is inherited. Height is controlled by genetic factors. The more closely related two people are, the more similar their height. It is also true that Indian children living in the rural areas of most Central or South American countries are significantly shorter than the Indian children living in the urban areas of those countries. Jensen's logic would suggest that the shorter stature of the rural children is due to a different genetic constitution. However, the data indicate otherwise. The shorter heights of the rural children do not seem to be due to heredity but to disease and environmental malnutrition. The heights of children in many areas of the world, including the United States, have increased considerably during the past twenty years due to better nutrition and immunization against disease, not as a result of changes in genetic structure. Yet a person's height is still subject to genetic control. The essential error in Jensen's argument is the conclusion that if a trait is under genetic control, differences between two populations on that trait must be due to genetic factors. This is the heart of Jensen's position, and it is not persuasive.

Professor I. I. Gottesman, a leading behavioral geneticist, also questions the validity of Jensen's ideas. He notes that, ". . . even when gene pools are known to be matched, appreciable differences in mean IQ can be observed that could only have been associated with environmental differences." In a study of 38 pairs of identical twins reared in *different environments,* the average difference in IQ for these identical twins was 14 points, and at least one quarter of the identical pairs of twins reared in different environments had differences in IQ score *that were larger than 16 points.* This difference is larger than the average difference between black and white populations. Gottesman concludes, "The differences observed so far between whites and Negroes can hardly be accepted as sufficient evidence that with respect to intelligence the Negro American is genetically less endowed."

Let us consider some additional empirical evidence that casts doubt on the validity of Jensen's position. Longitudinal studies being conducted in our laboratory reveal that lower class white children perform less well than middle class children on tests related to those used in intelligence tests. These class differences with white populations occur as early as one to two years of age. Detailed observations of the mother child interaction in the homes of these children indicate that the lower class children do not experience the quality of parent-child interaction that occurs in the middle class homes. Specifically, the lower class mothers spend less time in face to face mutual vocalization and smiling with their infants; they do not reward the child's maturational progress, and they do not enter into

long periods of play with the child. Our theory of mental development suggests that specific absence of these experiences will retard mental growth and will lead to lower intelligence test scores. The most likely determinants of the black child's lower IQ score are his experiences during the first five years of life. These experiences lead the young black child to do poorly on IQ tests in part because he does not appreciate the nature of a problem.

A recent study of urban black children showed that the IQ distribution had two peaks. There was a large proportion of children with IQ scores around 60 and a much larger group whose distribution was normal and similar to that of white populations. The examiners felt that the very low IQ scores were a product of failure to understand the problem; failure to know what to do; failure to appreciate a test was being administered. This argument finds support in a recent study by Dr. Francis Palmer of the City University of New York. Dr. Palmer administered mental tests to middle and lower class black children from Harlem. However, each examiner was instructed not to begin any testing with any child until she felt that the child was completely relaxed, and understood what was required of him. Many children had five, six and even seven hours of rapport sessions with the examiner before any questions were administered. Few psychological studies have ever devoted this much care to establishing rapport with the child. Dr. Palmer found very few significant differences in mental ability between the lower and middle class populations. This is one of the first times such a finding has been reported and it seems due, in part, to the great care taken to insure that the child comprehended the nature of the test questions and felt at ease with the examiner.

We can quickly dismiss Jensen's suggestion that compensatory education is not likely to help black children. The value of Head Start or similar remedial programs has not yet been adequately assessed. It is not reasonable to assume that compensatory education has failed merely because eight weeks of a Head Start program organized on a crash basis failed to produce stable increases in IQ score. The flaws in this logic are overwhelming. It would be nonsense to assume that feeding animal protein to a seriously malnourished child for three days would lead to a permanent increase in his weight and height, if after 72 hours of steak and eggs he was sent back to his malnourished environment. It *may be* that compensatory education is of little value, but this idea has not been tested in any adequate way up to now.

Finally, it is important to realize that the genetic constitution of a population does not produce a specific level of mental ability; rather it sets a range of mental

ability. Thus genetic factors are likely to be most predictive of proficiency in mental talents that are extremely difficult to attain, such as creative genius in mathematics or music, not relatively easy skills. Learning to read, write or add are easy skills, well within the competence of all children who do not have serious brain damage. Therefore, it is erroneous to suggest that genetic differences between human populations could be responsible for failure to master school related tasks. Ninety out of every 100 children, black, yellow or white, are capable of adequate mastery of the intellectual requirements of our schools. Let us concentrate on the conditions that will allow this latent competence to be actualized with maximal ease.

References

Gottesman, I. I. Biogenetics of race and class. In M. Deutsch, I. Katz, and A. R. Jensen (Eds.), *Social class, race, and psychological development.* New York: Holt, Rinehart & Winston, 1968.

Palmer, F., unpublished research reported at a colloquium at Harvard University, November, 1968.

Has Compensatory Education Failed? Has It Been Attempted?

J. McV. HUNT, *University of Illinois*

While Professor Hunt finds much of interest in parts of Jensen's article, he objects strongly to some of its conclusions. Hunt fails to find satisfactory evidence that we may make the assertions about genetic differences determining the intelligence of Negroes and whites which Jensen has offered. He finds Jensen's claims about the high heritability of intelligence unsubstantiated; he finds Jensen's conclusion that observed group mean differences in IQ scores among Negro and white populations are genetically determined to be even less supportable. Hunt offers an alternative hypothesis: given the necessary relationship between the physical structure of the nervous system and the behavior of the system (as in IQ), we must provide rich post-natal experience in order to develop the inherent structures. He offers analogies from animal research which suggest that the physical development of the brain is directly influenced by its information-processing activities—these activities are particularly effective in neo-natal organisms.

Jensen's paper is a critical effort to correct the currently wide-spread "belief in the almost indefinite plasticity of intellect." He asserts that "the ostrich-like denial of biological factors in individual differences, and the slighting of the role of genetics in the study of intelligence can only hinder investigation and understanding of the conditions, processes, and limits through which the social environment influences human behavior" (p. 29). He finds my term "fixed intelligence" to be rather misleading for two real and separate reasons: (1) the genetic basis of individual differences in intelligence and (2) the stability or the constancy of the

IQ throughout the individual's life. A major share of his paper is devoted to explaining the heritability of traits and to the theoretical and empirical basis for the proposition that about 80% of the individual variance in intelligence (defined in terms of the IQ and/or Spearman's *g*) has a genetic basis. This, at least by implication, explains why compensatory education "apparently has failed" (p. 2). He examines class differences and race differences in these same terms. But there is more to his paper. In the end, he offers, from the results of his own investigations, a basis for some hope through education if educational practice is modified.

Honest criticism is useful, both in science and in the process of social change which the behavioral, biological, and social sciences have now begun to influence. It is always useful unless it serves to hamper freedom of and support for investigation and for the development of appropriate technologies for coping with social problems. On the whole, Jensen's criticism comes in a constructive spirit. Moreover, it is informative. I am glad for the invitation to respond to his paper, for it has motivated more careful reading and consideration than I might otherwise have given it. In responding, I would like to synopsize his argument and respond point by point, but in the pages allowed me, I must respond selectively.

It is worth noting that Professor Jensen's argument is highly sophisticated in terms of both psychometrics and population genetics. His explanations in these domains are as briefly clear and as uncluttered with unnecessary jargon as any I have seen. He defines intelligence operationally in terms of what the IQ tests measure, of what accounts for the co-variation among test scores, (Spearman's *g*), and of the relations of these measures to scholastic ability (whence the tests come originally in the work of Binet and Simon), to occupational status, and to job success. What the IQ measures and what Spearman's *g* represents psychologically, he writes, "is probably best thought of as a capacity for abstract reasoning and problem-solving ability" (p. 19), and is also epitomized in cross-modality transfer. He recognizes clearly that intelligence is a phenotype, not a genotype:

> . . . the IQ is not constant, but, like all other developmental characteristics, is quite variable early in life and becomes increasingly stable throughout childhood. By age 4 or 5, the IQ correlates about .70 with IQ at age 17, which means that approximately half [r^2] of the variance in adult intelligence can be predicted as early as age 4 or 5. (p. 18)

He does not note here that this increasing stability is based on a part-whole relationship wherein the IQs of successive ages constitute increasing proportions of IQ of the criterion age. He asks the traditional geneticists' question of how much

variation (i.e., individual difference) in measures of the intelligence phenotype of our population can be accounted for in terms of variation in genetic factors. He then presents the evidence for heritability (H) approximating .80 in European and North American Caucasian populations. Jensen explicitly accepts that the value of H holds only for the population sampled, and that under changed conditions the value of H could be expected to change.

Despite this psychometric and genetic sophistication in Professor Jensen's discourse, I find little evidence of an inclination to broaden the nomological net to include evidence from social psychology, from the physiological effects of early experience in animals, and from history to help interpret the psychometric and genetic findings. I find wanting an appreciation of how what Sumner (1906) called the "folkways" and Sherif (1936) has called the "social norms" can operate to produce radically different ecological niches for developing infants and children of differing social classes and races. I find wanting also an appreciation of individual lives as dynamic processes in which the preprogrammed information in the genetic code get cumulatively modified in both rate and direction by successive adaptations to the circumstances of the ecological niche. Thus, Professor Jensen's argument sums up to a sophisticated justification of what I have termed, and perhaps unfortunately, "fixed intelligence" and "predetermined development" (Hunt, 1961). Except for the educational significance he finds in the results of his own investigations, his argument allows only a eugenic approach to the problems of incompetence and poverty. With the exception of this loophole it is a counsel of despair, for our increasingly technological society cannot afford a century or two of selective breeding.

Points of Agreement

Even though my own theoretical predilections (or prejudices, perhaps) differ sharply from those of Professor Jensen, I have found many points in his paper with which I agree heartily. We agree, albeit for different reasons, that the concept of the "average child" is highly unfortunate in education. I find myself delighted with his thumbnail sketch of the central features of that traditional educational practice which has consequently evolved in Europe and America. It is the best I have ever seen. Unlike Jensen, however, I do not find imagining radically different forms so difficult even though I recognize that changing our educational folkways will be exceedingly difficult.

I agree that there is abundant evidence of genetic influences on behavior and

that one can increase or decrease by selective breeding the measures of any phenotypic trait which has been investigated, but I believe from evidence omitted in Jensen's discourse that what Dobzhansky has termed the "range of reaction" (Sinott, Dunn, & Dobzhansky, 1958, p. 22ff) is probably greater for intelligence than it is for many other characteristics which depend less on what I suspect are cumulative effects of successive adaptations.

I agree that it is essentially meaningless to speak of "culture free" and of "culture fair" tests, and yet I also agree that Cattell (1963) has made, on the basis of differences within the intercorrelations, "a conceptually valid distinction between two aspects of intelligence, *fluid* and *crystallized*" (p. 13).

I agree with Jensen that the technological advances in our culture make it highly important to raise the intelligence, the educational attainments, and/or the general competence of those people who now comprise the bottom quarter of our population in measures of this cluster of characteristics. I agree that the national welfare policies we established in the 1930s have probably operated in disgenic fashion, and that it is highly important to establish welfare policies which will encourage initiative and probably, in consequence, help foster positive genotypic selection.

I could not agree more completely than I do with Professor Jensen's statement that:

> The variables of social class, race, and national origin are correlated so imperfectly with any of the valid criteria on which [social] decisions [with respect to individuals] should depend, or, for that matter, with any behavioral characteristic, that these background factors are irrelevant as a basis for dealing with individuals—as students, as employees, as neighbors. (p. 78)

Finally, for me the most interesting portion of Professor Jensen's paper is to be found in the results of his own investigations. The absence of class differences in what he calls "associative" learning, despite substantial differences in "cognitive" learning, is exceedingly interesting. Although I may well give a quite different interpretation of the basis for these findings than does Professor Jensen, I agree equally strongly with the educational implication he draws from his findings. *One does not provide equality of educational opportunity by submitting all children to the lock-step and by providing them with a single way in which to develop their genotypic potential.* Variation in genotypes combines with variation in early experience to call for an increased individualization of education. (Jensen's discussion is on pp. 6-8, 111-117.)

Points of Disagreement

Although I have found many points in Jensen's paper with which I can heartily agree, I have also found others with which I can just as heartily disagree. These are, first, several matters concerned with the measurement, the distribution, the development, and the nature of intelligence; second, the nature of his emphasis on biological versus psychological and social factors in behavioral development and the implications he draws for the relatively fixed nature of the existing norms for "intelligence." Third is Jensen's implicitly limited view of the learning process, coupled with his apparent lack of appreciation of the cumulative and dynamic implications of existing evidence of plasticity in the rate of behavioral development. Fourth are the implications which he draws for class and race differences from the measures of heritability of the IQ in European and American Caucasians. Finally, comes a disagreement about the wisdom of his opening sentence that "compensatory education has been tried and it apparently has failed" in the light of his avowed predilection for keeping all hypotheses open to investigation (and hopefully to technological development) as well as debate.

Matters Concerned with Intelligence

First, I find definitions of intelligence in terms of existing psychometric operations highly unsatisfying. Even though it was J. P. Guilford who introduced me to psychology and attracted me to the field largely with his discourse on aptitude testing and its implications for vocational guidance, I must confess that I have long distrusted the statistical operations of correlational analysis and averaging once they leave me without at least an intuitive connection with behavioral and biological observables. Thus, when Jensen remarks that Spearman's *g*-factor has "stood like a rock of Gibraltar," I find it hard to take seriously his avowance that "we should not reify *g* as an entity, of course, since it is only a hypothetical construct intended to explain covariation among tests" (p. 9). The *g*-factor explains on the average some 50% of the total variation in individual differences. Jensen notes further that "as the tests change, the nature of *g* will also change, and a test which is loaded, say, .50 on *g* when factor analyzed among one set of tests may have a loading of .2% or .8%, or some other value, when factor analyzed among other sets of tests" (p. 11). Apparently *g* is the most malleable and ameoboid rock extant. Jensen, however, makes a partial escape from his self-made operational *cul-de-sac* by arguing that intelligence is but one component of ability and competence. Thus, his own investigative finding that children of lower-class back-

ground can manage "associative" learning as well as children of middle-class background provides him with a ray of educational hope.

Professor Jensen devotes a substantial portion of his paper to an explication of the existing distribution of IQs in the population. He makes much of the basic normality of the distribution and the deviations from normality for pathological retardates and the "bulge" between 70 and 90 which he attributes to "the combined effects of severe environmental disadvantages and of emotional disturbances that depress test scores" (p. 27). Professor Jensen acknowledges that the traditional procedures provided by Binet and Simon for determining the mental age of any test-item forces the scores to assume a normal distribution, and he honestly admits that "the argument about the distribution of intelligence thus appears to be circular" (p. 21). He then argues that the only way out is to look for evidence that intelligence scales behave like an "interval scale." He finds the most compelling evidence from studies of the inheritance of intelligence. Am I emitting a mere flippancy if I respond that apparently, for Jensen, going twice around the circular argument removes its circularity? Actually, I find no serious fault with this discussion of the existing distribution of IQs in the population until Jensen begins to draw from it the implication that this existing distribution is fixed in human nature for all time, or until selective breeding alters it. My reasons for finding fault with this implication are derived from enlarging the nomological net to include evidence from outside the domains of psychometrics and population genetics as applied to intelligence, and I hope my argument will gradually become both clear and forceful.

On the matter of the stability of the IQ, Professor Jensen disavows any claim for constancy. On the other hand, he appears to view intellectual development as a matter of static, largely predetermined, growth. Thus he takes the findings of Bloom (1964) and emphasizes that half of the variance in the IQ at age seventeen can be predicted from IQs at ages of four and five years. If one considers the development of intelligence to be in substantial degree a function of the cumulative effects of informational and intentional interaction with physical and social circumstances, and if one takes into account the fact that the longitudinal predictive value of the IQ involves part-whole relationships, the emphasis can readily be reversed. Thus, just as embryologists have said that half of the epigenetic changes in a human life occur between conception and the end of the embryonic phase after only two months of gestation, it is more than a mere analogy to say that half of the epigenetic changes in mental development have typically taken place by about age four. This latter position puts the emphasis on the importance of

early experience (including the intrauterine and nutritional) as both Bloom and I have been wont to do.

Perhaps I am wrong in inferring that Professor Jensen at least implicitly conceives a sharp distinction between tests of intelligence and tests of educational achievement, for he emphasizes that the former has substantially a higher heritability (80%) than the latter (approximately 60%). Because the main thrust of his paper is to emphasize the high heritability of intelligence, one can understand his omission of the papers by both Ferguson (1954, 1956, 1959) on the relation of learning to human ability and Humphreys (1962a, 1962b)[1] on the point that tests of intelligence and tests of academic achievement differ only in degree, in the sense that the former assess the results of incidental learning typically distant in time from that of the testing while the latter assess the results of learning in specific educational situations near in time to the testing. When one combines the evidence and arguments from these papers with a conception of intelligence as a cumulative, dynamic product of the ongoing informational and intentional interaction of infants and young children with their physical and social circumstances, one must call into question the notion of intellectual development as essentially a static function of growth, largely predetermined in rate.

The Dualism of Biological Versus Psychological (and Social) Factors

Professor Jensen quotes with high approval a paragraph by Edward Zigler to the effect that: "Not only do I insist that we take the biological integrity of the organism seriously, but it is also my considered opinion that our nation has more to fear from unbridled environmentalists than from those who point to such integrity as one factor in the determination of development.... It is the environmentalists who have placed on the defensive any thinker who, perhaps impressed by the revolution in biological thought stemming from discoveries involving DNA-RNA phenomena, has had the temerity to suggest that certain behaviors may be in part the product of read-out mechanisms residing within the programmed organism" (p. 29).

[1] Professor Hunt calls attention to research that was omitted in the pre-publication draft on which this discussion is based. The Humpreys data is included in the printed version of Jensen's article as a note on page 58. The reader's attention is directed to the opposite interpretations each author draws from the research. In effect, Hunt argues that the correlation of IQ and academic achievement indicates that IQ is dynamic and cumulative; Jensen holds to his conception of IQ as largely predetermined, and suspects that he has overestimated the malleability of academic achievement.

Has Compensatory Education Failed?
J. MCV. HUNT

I believe that I have regularly taken "the biological integrity of the organism" seriously. Taking seriously the biological integrity of the organism is the major reason for my repeated concern with what I call "the problem of the match" between what has been built into the organism—through the program of maturation and through previous informational interaction with circumstances—and how newly encountered circumstances affect his motivation and continuing development (see Hunt, 1961, pp. 268-288; 1965; 1966, pp. 118-132). Also motivated by serious concern for the biological integrity of the organism is an extended effort to develop sequential ordinal scales of psychological development (Uzgiris & Hunt, 1969) and to look toward what one might term a "natural curriculum" for the fostering of early psychological development. In addition to these remarks, which may be regarded as defensive, it may be worth noting that the RNA (ribonucleic acid) phenomena are chiefly products of an organism's adaptation to circumstances.

Throughout his paper, and especially when he comes to the section on "how the environment works," the thrust of Professor Jensen's argument is to place psychological factors (and the social subset of these factors) in a kind of dualistic opposition to biological factors. Having implicitly constructed the dualism, he proceeds to denigrate the importance of the psychological set relative to the importance of biological set.

First, let me dispose of the dualism. Ample evidence has now accumulated to show that the consequences of informational interaction with circumstances, through the ears and the eyes (and especially the latter for the evidence extant), is quite as biological in nature as the effects of nutrition or of genetic constitution. Interaction through the eyes, especially early in life, has genuine neuroanatomical and neurochemical consequences.

Much of this evidence has its conceptual origin in the theorizing of Donald Hebb (1949). It was Hebb's hypothesis that the development of form-vision derives from sensory (S-S) integration that prompted Riesen and his colleagues to rear chimpanzees in the dark in order to determine the effect of light stimulation on the function and structure of the visual system. As is now widely known, a period of 16 or 18 months in total darkness produced drastic effects. On the functional side, there were a number of defects which proved essentially irreversible in those chimpanzees submitted to total darkness for 16 months or longer (see Riesen, 1958). On the side of anatomical structure, a defect was manifest during life as a pallor of the optic disc (Riesen, 1958). When these animals were sacrificed after some six years in full daylight, a histological examination brought out

clear evidence of defects in the ganglion-cell layer of the retinae and in the optic nerve. These anatomical consequences within the visual system had themselves been irreversible (Chow, Riesen, & Newell, 1957). The histological examination also got evidence of a paucity of Mueller fibers within the retinal ganglia, and it should be noted that Mueller fibers are glia (Rasch, Swift, Riesen, & Chow, 1961).

Another line of investigation has stemmed from Hydén's (1961) biochemical hypothesis that memory and learning involve the metabolism of ribonucleic acid (RNA) in an interaction between neural and glial cells of the retina and brain. Hydén's hypothesis prompted Brattgård (1952) to rear rabbits in the dark. Histochemical analysis of the retinae of these dark-reared rabbits revealed a deficiency in RNA production of their retinal ganglion cells as compared with their light-reared litter-mates. Since then histological and histochemical effects of dark-rearing have been found not only in chimpanzees (Chow, *et al.*, 1957) and rabbits, but also in kittens (Weiskrantz, 1958) and in rats (Liberman, 1962).

I have often expressed the wish that someone would extend this line of investigation centrally in the visual system to the lateral geniculate body of the thalamus and to the striate area of the occipital lobe. After regaling Robert Reichler of the National Institute of Mental Health with this evidence just outlined, I expressed again this wish to see an extension to the lateral geniculate body and to the striate area of the occipital lobe. Dr. Reichler responded excitedly that this had been done. In late October, he had attended an NIMH-supported conference on dyslexia where Dr. F. Valverde of Cajal's Institute in Madrid had presented a paper authored with Ruiz-Marcos which indeed reported such investigations with highly interesting findings. I am indebted to Dr. Reichler for letting me see a copy of the conference draft of the paper by Valverde and Ruiz-Marcos.

As yet I have had no opportunity to examine the evidence in detail, but their paper reviews an investigation by Wiesel and Hubel (1963), in which were described clearly evident defects in the cell areas of the lateral geniculate bodies on the thalami of kittens corresponding to the single eye deprived of vision for three months. Their paper also reviews evidence from investigations by Gyllesten (1959), by Coleman and Riesen (1968), by Ruiz-Marcos and Valverde (1968), by Valverde (1967, 1968), and by Valverde and Esteban (1968). All these investigations have shown clearly the effects of being reared in the dark, sometimes for only a very few days, on the fine structure of the striate area of the occipital lobe which is the center for visual reception. These effects show in the dendritic fields, and they show especially as a diminution in the number of spines on the dendrites of the large pyramidal cells in the striate area of the visual cortex (Val-

verde, 1967, 1968). Through electron-microscopy it was determined that the number of spines on these dendrites, in intervals at given distances from the wall of the cell body, is ordinarily very highly correlated with mouse age, but when mice are reared for various periods in the dark, this correlation is markedly diminished (Ruiz-Marcos & Valverde, 1968), and the diminution is especially marked for the days immediately after the eyes open. Clearly the psychological factor of dark-rearing produces neuro-anatomical and neurochemical effects not only in the eye but in the thalamus and in the visual area of the cortex. Thus, this psychological factor of visual function appears to be quite as biological in its consequences as are the consequences of nutrition and genotype.

Dark-rearing produces just the kind of anatomical effects one might envisage from Hebb's (1949) concepts of "cell assemblies" and "phase sequences." I see no reason to think that such processes should be less likely in human beings than in rodents. It takes little imagination, moreover, to extrapolate from these findings. I suspect that sensorimotor functioning, especially during the earliest phases of behavioral development in the first and second years, influences the development of such things as the spines on dendrites throughout the brain. The success of Hydén and Egyhazi (1962) in identifying with remarkable specificity the locus of the neuroanatomical and neurochemical effects of rats learning to climb a guy-rope suggests that each coordination, between vision-and-hand motion or between eye-function and ear-function, has its own neuro-electrical-chemical-anatomical equipment. I suspect that when such equipment has emerged as the consequence of a given bit of functional accommodation or learning, it can readily be employed in other functioning and thereby become the basis for the transfer of training. Moreover, as equipment has been developed in many domains, it can in all likelihood become one of the bases for the positive intercorrelation among tested abilities which Spearman called *g*.

In his section on "how the environment works" Professor Jensen contends that "below a certain threshold of environmental adequacy, deprivation can have a markedly depressing effect upon intelligence. But above this threshold, environmental variations cause relatively small differences in intelligence." He contends further: "The fact that the vast majority of the populations sampled in studies of the heritability of intelligence are above this threshold level of environmental adequacy accounts for the high values of the heritability estimates and the relatively small proportion of IQ variance attributable to environmental influences" (p. 60). The evidence of increase in the development of brain structures following enrichments of early experience are hardly consonant with this position. Alt-

man and Das (1964), for instance, have reported a higher rate of multiplication of glial cells in the cerebral cortices of rats reared in "enriched environments" and in rats reared in the "impoverished environments" of laboratory cages. In another extended program of such investigation which has been underway for more than a decade at the University of California, Bennett, Diamond, Krech, and Rosenzweig (1964) and Krech, Rosenzweig, and Bennett (1966) have done a long series of studies which indicate that rats reared in relatively complex environments have shown cortical tissue greater in weight and thickness than that of litter-mates reared in the simpler environments of laboratory cages. Here "complexity" has been defined in terms of the variety of objects available for the rats to perceive and to manipulate and the variety of different kinds of space to be explored. These rats reared in complex environments have also shown histochemical effects in the form of higher total acetylcholinesterase activity of the cortex than the cage-reared rats. Associated with these neuroanatomical and neurochemical effects of the life history, moreover, is a higher level of maze-problem-solving ability in the rats reared under complex circumstances than in those reared in laboratory cages.

The definition "of a certain threshold of environmental adequacy" is unclear, but it can be said that cage-rearing is the standard ecological niche of laboratory rats and that it involves no serious absence of light and sound. Contrary to Jensen's position that it is only below "a certain threshold of environmental adequacy" that there can be a markedly depressing effect on intelligence, I am inclined to suspect that the basic central equipment for the inter-modal transfer which Jensen conceives to be a prime example of Spearman's *g* can be greatly modified by the informational interaction of the human infant and young child with his physical and social circumstances. I say that I suspect this is the state of affairs. This statement has not been proven, but the thrust of the existing evidence points strongly in the direction which I have indicated.

Learning and the Cumulative Implication of Plasticity in Early Development

The traditional view of heredity and environment held them to be essentially separate processes in development, and maturation was conceived to be the developmental representative of heredity, with learning the developmental representative of environment. We have just seen that the young organism's adapta-

tions to the environment influence maturation, but we have not clarified the nature of learning.

Learning is typically conceived in terms of the ways it has been investigated in the laboratory. Investigations of learning still bear the marks of the pioneers: Ebbinghaus for rote learning, Bryan and Harter for skill learning, Pavlov for classical conditioning, and, for the fourth general category, C. Lloyd Morgan and E. L. Thorndike for trial-and-error with reinforcement, Clark L. Hull for instrumental learning motivated by drive and reinforced by drive-reduction, and B. F. Skinner for operant conditioning. If one examines the developmental observations of Piaget (1936, 1937), wherein accommodation and assimilation become the terms for learning, one finds several kinds of effects of encounters with circumstances which have failed to get investigated in psychological laboratories. If one examines the almost forgotten work on attention, the work of the ethologists, and the work of social psychologists on attitude change and communication, one finds other kinds of modification of function, and presumably of neuroanatomy and neurochemistry, through encounters with informational circumstances which do not get into the chapters on learning. I believe I have identified eight kinds of learning seldom studied for themselves which appear to be operative in psychological development (Hunt, 1966). The number is unimportant; the point is that Professor Jensen's distinction between associative learning and cognitive learning is but a conceptual drop in the bucket. His finding that the class-differences evident for cognitive learning are not evident for associative learning is exceedingly interesting, however.

What appears to be wrong with Professor Jensen's implicit conception of learning is that it consists only (or basically) of those minor changes of function which can be effected within short intervals of time in the laboratory. Thus, he speaks of learning ability as a kind of static trait which accounts for the number of trials required for the assimilation or mastery of relatively miniscule accommodations.

Except for the case where he calls for studies of the transfer of learning before age five to the cognitive functions after age six (in which I join him), I miss in his discourse any strong appreciation of what must be the cumulative dynamic effects of adaptations at one phase of development on the adaptations of later phases. Thus, he can write of the influence of the genotype "reading through the environmental overlay."

Although Professor Jensen acknowledges that such "extreme sensory and motor restrictions in environments such as those described by Skeels and Dye (1939) and Davis (1947), in which the subjects had little sensory stimulation of any kind

77-943 O - 72 - 39

and little contact with adults" (p. 60) resulted in large deficiencies in IQ, he tends to minimize their importance. He notes in favor of his view that the orphanage children of Skeels and Dye gained in IQ from an average of 64 at 19 months of age to 96 at age six as a result of being given "social stimulation and placement in good homes at between two and three years of age" (p. 60). He notes that when these children were followed up as adults, they were found to be average citizens in their communities, and their own children had an average IQ of 105 and were doing satisfactorily in school. Similarly, Davis (1947) reported the more extreme case of Isabel, who had an IQ of 30 at age six, but who, when put into an intensive educational program at age six, developed a normal IQ by age eight. From these examples, he contends that even extreme environmental deprivation need not permanently result in below-average intelligence.

Professor Jensen neglects to report the results of the follow-up study of the adult status of the Skeels-Dye children left in the orphanage (Skeels, 1966). Those who were removed from the orphanage before they were 30 months old and placed on a women's ward at a state institution for the mentally retarded, and then later adopted, were all self-supporting and none became a ward of any institution. Their median educational attainment was 12th grade. Four had one or more years of college work, one received a bachelor's degree and went on to graduate school. On the other hand, of the 12 children who remained in the orphanage, one died in adolescence following continued residence in a state institution for the mentally retarded, and four remained on the wards of such institutions. With one exception, those employed were marginally employed, and only two had married. It is true that the effects of early experience can be reversed; the point to be made here, however, is that the longer any species of organism remains under any given kind of circumstances, the harder it is to change the direction of the effects of adaptation to those circumstances.

Even in infants reared in middle-class homes evidence exists of a remarkable degree of plasticity in early behavioral development. In my own laboratory, for instance, Greenberg, Uzgiris, and Hunt (1968) have shown that putting an attractive pattern over the cribs of such infants beginning when they are five weeks old, reduces the age at which the blink-response becomes regular for a target-drop of 11.5 inches from a mean of 10.4 weeks, in infants whose mothers agreed to put nothing over the cribs of their infants for 13 weeks, to a mean of 7 weeks. In the familiar terms of the IQ ratio this represents an increase of 48 points for the blink-response. The differences between the groups in mean age for drops of 7 inches and for drops of 3 inches becomes progressively less. Thus, the findings are

quite consonant with those studies of the twenties and thirties which found the effects of practice on given skills to be evanescent. On the other hand, if one provides circumstances which permit the hastened looking schema, indicated by the blink-response, to be incorporated into a more complex sensorimotor organization, its early availability should be reflected in increased advancement. This is precisely the sort of thing one finds in the work of White and Held (1966). In their work, the capacity for visual accommodation permits looking to become incorporated into eye-hand coordination. In a normative study of successive forms of eye-hand coordination, top-level reaching failed to appear until the median age of the group was 145 days. The second enrichment program reduced the median age for top-level reaching from this 145 days to 87 days—an advance of 66 points in the familiar terms of the IQ ratio for this final level of eye-hand coordination. Hypothetically, at least, one should be able to extrapolate on this principle, but as yet experimental evidence is unavailable to confirm the hypothesis.

Cumulative and dynamic implications of this existing evidence of plasticity in the rate of behavioral development raises the question of what Dobzhansky has termed the "norm of reaction" (see Sinnott, Dunn, & Dobzhansky, 1958, p. 22ff) for the case of human intelligence. Although no one can now say how large the cumulative modifications in measurements of human intelligence might possibly be, Wayne Dennis (1966) has published a study which is highly relevant. The study examines the mean IQs from the Draw-a-Man Test for groups of typical children aged six and seven years from some 50 cultures over the world. Florence Goodenough (1926) devised this test to be culture free. Its freedom from cultural influences was called into question, however, when typical Hopi Indian children of six and seven turned up with a mean of 124 on the test (Dennis, 1942). This mean of 124 equaled the mean IQ for samples of upper-middle-class suburban American children and for samples of children from Japanese fishing villages. The lower end of this distribution of mean IQs finds nomadic Bedouin Arab children of Syria with a mean IQ of 52. Here, then, we find direct evidence of a norm of reaction of about 70 points in Draw-a-Man IQ. The most obvious correlate of this variation in mean IQ is amount of contact with the pictorial art. Among Moslem Arab children, whose religion prohibits representative art as graven images, the range in mean Draw-a-Man IQ is from 52 to 94, and the most obvious correlate of this norm of reaction is contact with groups of the Western culture. This is the most direct evidence concerning the norm of reaction for human intelligence of which I know. While the factor structure of the Draw-a-Man Test is probably considerably less complex than is that of either the Stanford-Binet

or the Weschler Children's Scale, within our own culture Draw-a-Man scores correlate about as well with those from these other more complex scales as scores on them to with each other.

In connection with this discussion of the norm of reaction, which Professor Jensen mentions but to which he gives little attention, it is interesting to note what he omits from a paragraph quoted from the geneticist, Dobzhansky (1968b, p. 554 quoted in Jensen, p. 30). The omitted portion reads: "Although the genetically-guaranteed educability of our species makes most individuals trainable for most occupations, it is highly probable that individuals have more genetic adaptability to some occupations than to others. Although almost everybody could become, if properly brought up and properly trained, a fairly competent farmer, or a craftsman of some sort, or a soldier, sailor, tradesman, teacher, or priest, certain ones would be more easily trainable to be soldiers and others to be teachers, for instance. It is even more probable that only a relatively few individuals would have the genetic wherewithal for certain highly specialized professions, such as musician, or singer, or poet, or high achievement in sports or wisdom or leadership."

Finally, I am among those few who are inclined to believe that mankind has not yet developed and deployed a form of early childhood education (from birth to age five) which permits him to achieve his full genotypic potential. Those studies which so sharply disconfirmed what R. B. Cattell (1937) once characterized as a "galloping plunge toward intellectual bankruptcy," (see Hunt, 1961, p. 337ff) can probably be repeated again after 20 to 25 years if our society supports the necessary research and development of educational technology to enable us to do early childhood education properly. In connection with this possibility of a general increase in intelligence, we should consider also what has happened to the stature of human beings. It appears to have increased by nearly a foot without benefit of selective breeding or natural selection. While visiting Festival Park in Jamestown, Virginia recently, we examined the reproductions of the ships which brought the settlers from England. They were astoundingly small. The guide reported that the average height of those immigrants was less than 5 feet, and that the still famous Captain John Smith was considered to be unusually tall at 5 feet 2 inches. The guide's "instruction book" puts the authority for these statements in the Sween Library at William and Mary. I have been unable to check the evidence, but scrutiny of the armor on display in various museums in England implies that the stature of the aristocrats who wore it must typically have been about the reported size of those immigrants to Jamestown. Also, the guide

for the U. S. *Constitution* includes in his spiel the statement that the headroom between decks needed to be no more than 5 feet and 6 inches because the average stature of sailors in the War of 1812 was but 5 feet and 2 inches. This increase in height can occur within a single generation. Among the families of German Russians whom I knew while growing up in Nebraska it was typical to find the average height of the children several inches above mid-parent height, and I can cite instances in which the increase was approximately a foot where all the children were sons. Inasmuch as Professor Jensen resorts repeatedly to the analogy between intelligence and stature, such evidence of an increase in the average height for human beings, the reasons for which are still a matter largely of conjecture, should have some force in increasing the credibility for the genetic potentiality for a general increase in intelligence.

Implications from Existing Measures of Heritability

Professor Jensen recognizes explicitly that measures of heritability may change as the nature of the population changes. Nevertheless, from these existing measures of heritability in European and American Caucasians, he draws implications for both class and race differences which, in view of the considerations already presented, I simply cannot accept at face value.

From the physiological evidence, from the fact that one can readily hasten the development of sensorimotor organizations in children of the middle class, and from the fact that technological advances have quite regularly increased the mean IQ of populations, I see no reason to believe that the current distribution of intelligence is fixed by the biological nature of man, despite the fact that heritability studies indicate that approximately 80% of the individual variance in the IQ can be attributed to variations in genotypes. Moreover, in view of the sharp contrast between the child-rearing practices of the middle class with those of the people of poverty, I see no reason to believe that the class differences now evident are inevitable. Finally, inasmuch as black people have had more than a century in slavery and then, since the war between the States, another century in both poverty and the bondage of "folkways," I see no reason to consider existing race differences as inevitable.

The contrast between the child-rearing of the middle class and that of the poor needs to be better understood. A study by Maxine Schoggen at the Demonstration and Research Center for Early Education at the George Peabody College for Teachers in Nashville is bringing out this contrast more forcefully than any other

of which I know. The studies concern samples of eight families of professional status, eight of rural poverty, and eight of urban poverty. The families of rural poverty are white; those of urban poverty are black. In each family there is a 3-year-old who is the target-child. Observer-recorders become so well acquainted with these families that they become like furniture. They record for equal periods of time in functionally equivalent situations like meal time, bed time, and the time when the older children return from school. The observers record the instances of social interaction initiated by the older members of the family toward the target-child, and their reactions to the interaction initiated by the child. These are termed "environmental force units." From the evidence available, the older members of professional families initiated somewhat more than twice as many "environmental force units" per unit of time toward the 3-year-old in their family as did the older members in the families of either urban or rural poverty. I have asked Dr. Schoggen about how much difference there might be in the frequency of units in which the older members of the family would call upon a child to note the shape, the size, the color, or even the placement of objects and persons. She has indicated that this is quite common in professional families, but that it seldom occurs in the families of poverty except in connection with errands. Then the child usually gets castigated for his stupidity. On the verbal side, professional parents often call upon their three-year-olds to formulate such matters in language of their own, but families of either rural or urban poverty almost never do. One should recall in this connection that "warm democratic" rearing was associated with an average gain of 8 IQ points, over a three year period between the ages of approximately three or four to six or seven, in the study of Baldwin, Kalhorn, and Breese (1945), while the mean IQ dropped a point or two in the children of parents employing what these authors characterized as "passive-neglectful" and "actively-hostile" child-rearing (Baldwin, 1955, p. 523) This contrast between the rearing practices in families of professional status with those in families of either rural or urban poverty appears to be sharper than that between the families utilizing the various kinds of child-rearing identified by Baldwin, *et al.* Few if any of the studies of heritability have included the truly poor, so they have missed this portion of the variation in the circumstances of rearing.

At least a substantial portion of parents of poverty can be taught, however, to be effective teachers of their young when they are given models to imitate, when the actions of the models are explained, and when home visitors are provided to bring the new ways of child-rearing into the home (Gordon and associates, 1969; Karnes, 1969; Klaus & Gray, 1968; Miller, 1968). Moreover, when parents are in-

volved in the education of their young children, they communicate new-found practices to their neighbors and the parents themselves take a new lease on life. In the Karnes project, the mothers agreed that if they were to give each child the attention needed, they dare not have a new baby each year, and so they all enrolled at the local Planned Parenthood Clinic. Miller (1968) reports that in the extension of the Early Training Project a majority of the mothers have upgraded their skills, and the families in the projects have formed clubs—one in which husbands and wives bowl regularly.

It will be no easy matter to spread this kind of training to all the families of poverty throughout this country, but a start has been made in the Parent and Child Centers which the Office of Economic Opportunity has established on a pilot basis.

The enrolling in the Planned Parenthood clinic suggests that this kind of enterprise in early childhood education instigates help to prevent some of the disgenic processes with which Professor Jensen and I are both concerned.

I applaud Professor Jensen's proposal to develop a curriculum based upon his finding that children of lower-class background are equal in "associative" learning to children of middle-class background. In doing so, he may ultimately help to raise the general level of competence, and even the intelligence defined as Spearman's *g*, in the next generation of those who receive the benefit of his efforts to develop new educational technology. Moreover, since the effects of early experience can be reversed, at least in part, if and when Professor Jensen builds educationally upon the capacity of children from lower-class background for "associative" learning, he will probably increase measures of their *g*-factor gradually. His program will also probably increase measures of Cattell's "crystallized" intelligence in his pupils. To a lesser degree his program may also increase measures of their "fluid" intelligence. Moreover, Professor Jensen's program could well contribute to an increase in the intelligence of the next generation.

If one views societal evolution as a process, the mean of the IQ on the basis of existing standardizations and the existing measures of heritability can well be seen as the pre-measures to be compared with post-measures (based in the case of the IQ, of course, on today's standardizations) 10 or 20 years hence.

The Opening Sentence

At one point in his paper, Professor Jensen makes an ardent plea for keeping all hypotheses open for debate and investigation. With this plea, I heartily agree.

Unfortunately, since social change is a process, one cannot settle the issue between my reading of the broad range of evidence and his reading of contemporary evidence from existing distributions of IQs and contemporary measures of heritability, until these changes in the ecological niche of infants and young children, to be accomplished by the research, the development, and the deployment of early childhood education, have been available for at least a decade or two. Saying outright that "compensatory education has been tried and it apparently has failed" is but a half-truth. Moreover, it is but a half-truth which can help to boost the forces of reaction which could halt support for research on how to foster psychological development, for the development of technology of early childhood education, and for the deployment of that technology across the USA. Insofar as it succeeds in boosting these forces of reaction, it will leave the issue open only for debate. Once the support for investigation, development, and deployment has been removed, the differences between our readings of the evidence will no longer be open for "investigation.

Perhaps I should explain why Professor Jensen's sentence is but a half-truth. In this sentence, "compensatory education" implies Head Start, for it is Head Start which has been tried—at least a little. Project Head Start did deploy a form of early childhood education for which many had hopes of compensatory effects. It was hoped that giving children of the poor a summer or two or a year of nursery school, beginning at age four, would overcome the handicaps of their earlier rearing. I hoped it would, but I feared from the beginning that such broad deployment of a technology untested for the purpose might lead to an "oversell" which, with failure of the hopes, would produce an "overkill" in which would be lost, for who knows how long, the opportunity to bring into the process of social change, in the form of early childhood education, the implications of the various lines of evidence indicating the importance of early experience for intellectual development. The 1967 report of the U. S. Commission of Civil Rights is correct in stating that Head Start has not appreciably raised the educational achievement of the children who participated. There is, however, a reason which absolves compensatory education as such.

Maria Montessori in Italy and Margaret McMillan of England established nursery schools to aid the children of the poor. These were brought to America along with the intelligence tests and just as the emphasis on learning by doing was becoming established. Nursery schools did not survive in America as aids for children of the poor. Rather, they got adapted for the children of the well-to-do who could pay for them. Moreover, when the psychoanalytic movement coalesced

with Froebel's kindergarten movement and with the Child Study movement of G. Stanley Hall, the goal became one of freeing young children, for at least part of each day, from their mothers' strict disciplinary controls. Free play became the mode. Since such nursery schools constituted the only early education model available when Project Head Start began, traditional nursery school practice was the kind of early education deployed for the most part.

But Head Start is not synonymous with compensatory education. Professor Jensen knows this for he reviews a number of the investigations of compensation in one of the later sections of his paper. Compensatory education has not failed. Investigations of compensatory education have now shown that traditional play school has little to offer the children of the poor, but programs which made an effort to inculcate cognitive skills, language skills, and number skills, whether they be taught directly or incorporated into games, show fair success. A substantial portion of this success endures. If the parents are drawn into the process, the little evidence available suggests that the effect on the children, and on the parents as well, increases in both degree and duration. All this in seven years sounds to me like substantial success. Yet, we still have a long way to go before we shall have learned what an appropriate curriculum for infants from birth to five might be. Thus, Jensen's opening statement is a half-truth, and a dangerous half-truth, placed out of context for dramatic effect.

Insofar as the behavioral and educational sciences get involved in manning the tiller of social change, the practitioners of these sciences must learn to think in terms of processes and they must learn to think of political and social consequences of how and what they write and say. It does no good to plead for keeping all hypotheses open for debate and investigation if the form of the debate removes support for the relevant investigation and for the development and deployment required for a meaningful test of the hypotheses. I find it hard to forgive Professor Jensen for that half-truth placed out of context for dramatic effect at the beginning of his paper.

How much *can* we boost IQ and scholastic achievement by deliberately altering the ecological niche of infants and young children, from birth to age five, through early childhood education? Who knows? As I read the evidence, the odds are strong that we can boost both IQ and scholastic achievement substantially, but we cannot know how much for at least two decades. Moreover, we shall never find out if we destroy support for the investigation of how to foster early psychological development, for the development of educational technology, and for the deployment of that technology.

References

Altman, J. & Das, G. D. Autoradiographic examination of the effects of enriched environment on the rate of glial multiplication in the adult rat brain. *Nature*, 1964, **204**, 1161-1165.

Baldwin, A. L. *Behavior and development in childhood*. New York: Dryden Press, 1955.

Baldwin, A. L., Kalhorn, J. & Breese, F. H. Patterns of parent behavior. *Psychological Monographs*, 1945, **58**, No. 3, 1-75.

Bennett, E. L., Diamond, M. C., Krech, D., & Rosenzweig, M. R. Chemical and anatomical plasticity of the brain. *Science*, **146**, No. 3644 (1964), 610-619.

Bereiter, C., & Engelmann, S. *Teaching disadvantaged children in the preschool*. New York: Prentice Hall, 1966.

Bloom, B. S. *Stability and change in human characteristics*. New York: John Wiley & Sons, Inc., 1964.

Brattgård, S. O. The importance of adequate stimulation for the chemical composition of retinal ganglion cells during early post-natal development. *Acta Radiological* (Stockholm), 1952, Supplement 96, 1-80.

Cattell, R. B. *The fight for our national intelligence*. London: King, 1937.

Cattell, R. B. Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Education Psychology*, 1963, **54**, 1-22.

Chow, K. L., Riesen, A. H., & Newell, F. W. Degeneration of retinal ganglion cells in infant chimpanzees reared in darkness. *Journal of Comparative Neurology*, 1957, **107**, 27-42.

Coleman, P. D., & Riesen, A. H. Environmental effects on cortical dendritic fields. I. Rearing in the dark. *Journal of Anatomy* (London), 1968, **102**, 363-374.

Davis, K. Final note on a case of extreme isolation. *American Journal of Sociology*, 1947, **57**, 432-457.

Dennis, W. The performance of Hopi Indian children on the Goodenough Draw-a-Man Test. *Journal of Comparative Psychology*, 1942, **34**, 341-348.

Dobzhansky, T. On genetics, sociology, and politics. *Perspectives in Biology and Medicine*, 1968, **11**, 544-554.

Ferguson, G. A. Learning and human ability: A theoretical approach. *Factor analysis and related techniques in the study of learning*. Edited by P. H. DuBois, W. H. Manning, & C. J. Spies. A report of a conference held at Washington University in St. Louis, Missouri, February, 1959. Technical Report No. 7, Office of Naval Research Contract No. Nonr 816 (02).

Goodenough, F. L. *The measurement of intelligence by drawings*. Yonkers-on-Hudson, New York: World Book Co., 1926.

Gordon, I. J. (ed.) *Reaching the child through parent education: The Florida approach*. Gainesville: Institute for Development of Human Resources, College of Education, University of Florida, 1969.

Greenberg, D., Uzgiris, I., & Hunt, J. McV. Hastening the development of the blink-response with looking. *Journal of Genetic Psychology*, 1968, **113**, 167-176.

Gyllensten, L. Postnatal development of the visual cortex in darkness (mice). *Acta Morphologica* (Neerlando-Scandinavica), 1959, 2, 331-345.

Hebb, D. O. *The organization of behavior.* New York: John Wiley & Sons, Inc., 1949.

Humphreys, L. G. The nature and organization of human abilities. *The 19th Yearbook of the National Council on Measurement in Education.* Edited by M. Katz. Ames, Iowa, 1962a.

Humphreys, L. G. The organization of human abilities. *American Psychologist,* 1962b, 17, 475-483.

Hunt, J. McV. *Intelligence and experience.* New York: Ronald Press, 1961.

Hunt, J. McV. Intrinsic motivation and its role in psychological development. *Nebraska Symposium on Motivation,* 13. Edited by D. Levine. Lincoln: University of Nebraska Press, 1965.

Hunt, J. McV. Toward a theor [illegible] ed learning in development. In *Giving emphasis to guided learning.* Edited by [illegible] .I. Ojemann & K. Pritchett. Cleveland, O.: Educational Research Council, [illegible]

Hunt, J. McV. Poverty versus eq[illegible]ty of opportunity. In *The challenge of incompetence and poverty.* Urbana: Univ[illegible]ty of Illinois Press (in press).

Hydén, H. Biochemical aspects of brain activity. *Man and civilization: Control of the mind.* Edited by S. M. Farber & R. H. L. Wilson. New York: McGraw-Hill, 1961.

Hydén, H., & Egyhazi, E. Nuclear RNA changes of nerve cells during a learning experiment in rats. *Proceedings of the National Academy of Science,* 1962, 48, 1366-1373.

Karnes, M. B. *A new role for teachers: Involving the entire family in the education of preschool disadvantaged children.* Urbana: University of Illinois, College of Education, 1969.

Klaus, R. A., & Gray S. The early training project for disadvantaged children: A report after five years. *Monographs of the Society for Research in Child Development,* 1968, 33, No. 4.

Krech, D., Rosenzweig, M. R., & Bennett, E. L. Environmental impoverishment, social isolation, and changes in brain chemistry and anatomy. *Physiology and Behavior.* London: Pergamon Press, 1966.

Liberman, R. Retinal cholinesterase and glycolysis in rats raised in darkness. *Science,* 1962, 135, 372-373.

Miller, J. O. *Diffusion of intervention effects.* Urbana: ERIC Clearinghouse for Early Childhood Education, University of Illinois, 1968.

Rasch, E., Swift, H., Riesen, A. H., & Chow, K. L. Altered structure and composition of retinal cells in dark-reared mammals. *Experimental Cellular Research,* 1961, 25, 348-363.

Riesen, A. H. Plasticity of behavior: Psychological aspects. In H. F. Harlow & C. N. Woolsey (Eds.), *Biological and biochemical bases of behavior.* Madison: University of Wisconsin Press, 1958, 425-450.

Ruiz-Marcos, A., & Valverde, F. Mathematical model of the distribution of dendritic spines in the visual cortex of normal and dark-raised mice. *Journal of Comparative Neurology* (in press).

Sinnott, E. W., Dunn, L. C., & Dobzhansky, T. *Principles of genetics.* New York: McGraw-Hill, 1958.

Sherif, M. *The psychology of social norms.* New York: Harper, 1936.

Skeels, H. M. Adult status of children with contrasting early life experiences. *Monographs of the Society for Research in Child Development,* 1966, 31 (No. 3), 1-66.

Skeels, H .M., & Dye, H. B. A study of the effects of differential stimulation of mentally retarded children. *Proceedings of the American Association of Mental Deficiency,* 1939, 44, 114-136.

Sumner, W. G. *Folkways.* Boston: Ginn & Co., 1906 (1940).

Uzgiris, I. C., & Hunt, J. McV. *Toward ordinal scales of psychological development in infancy* (in press).

Valverde, F. Apical dendritic spines of the visual cortex and light deprivation in the mouse. *Experimental Brain Research,* 1967, 3, 337-352.

Valverde, F. Structural changes in the area striata of the mouse after enucleation. *Experimental Brain Research,* 1968, 5, 274-292.

Valverde, F., & Esteban, M. E. Peristriate cortex of mouse: Location and the effects of enucleation on the number of dendritic spines. *Brain Research,* 1968, 9, 145-148.

Weiskrantz, L. Sensory deprivation and the cat's optic nervous system. *Nature,* 1958, 181, 1047-1050.

White, B. L., & Held, R. Plasticity of sensorimotor development in the human infant. In J. F. Rosenblith & W. Allinsmith (Eds.), *The causes of behavior: Readings in child development and educational psychology* (2nd edition). Boston: Allyn & Bacon, 1966.

Wiesel, T. N., & Hubel, D. H. Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body. *Journal of Neurophysiology,* 1963, 26, 978-993.

Genetic Theories and Influences: Comments on the Value of Diversity

JAMES F. CROW, *University of Wisconsin*

Professor Crow agrees "for the most part with Jensen's analysis." He does suggest qualifications when drawing inferences from existing studies in biometrical genetics. First, he notes his reservations about the reality of the mathematical assumptions implicit in analysis of variance models. Second, he draws attention to the limited sample size available in studies of twins and siblings reared apart and asks how representative such groups are. Third, he notes that predictive models have inherent limits when new, qualitatively different, treatments are introduced into the environment.

Biometrical genetics has become quite a sophisticated subject with a substantial body of mathematical theory. One reason for this development is that the simple, mechanistic nature of Mendelian inheritance is very inviting to probabilistic model-builders. A second reason is that, because complex traits depend on more genes than can be individually identified, some sort of statistical treatment is necessary.

The foundations for the theory of biometrical genetics were laid by R. A. Fisher and Sewall Wright, using methods depending mainly on correlation and variance analysis. The procedures are widely used in animal breeding, thanks especially to J. L. Lush, who has been conspicuously successful in adapting these procedures to use in practical breeding problems. The general theory and methodology in this field are described with insight and lucidity by Falconer (1960).[1] Jensen's article, together with many others that he has written recently on this subject (see his bibliography), constitutes a thorough review and synthesis of

[1] This and subsequent references are to articles and books cited in Jensen's article.

Harvard Educational Review Vol. 39 No. 2 Spring 1969

the various attempts to apply these methods to human intelligence and scholastic achievement. Jensen has become a leader in this field, and I as a population geneticist, admi-e his understanding of the methods and his diligence and objectivity in bringing together evidence from diverse sources. He presents the evidence fairly, relying on empirical data in preference to introspection or traditional wisdom, and is very careful to distinguish between observation and speculation.

I shall confine my comments mainly to the genetic aspects of the article. I agree for the most part with Jensen's analysis. Any differences could probably best be described by saying that, in general, I have somewhat less confidence than he in the quantitative validity of the methods—more reservations about the reality of the necessary assumptions. I don't mean by this that I would reach opposite conclusions; I am simply more agnostic. This is especially true as regards intergroup comparisons and, in particular, the importance of genetic factors in racial differences.

The Concept of Heritability

Much of Jensen's article is concerned with the heritability of intelligence (I am trying to use the term intelligence in the same technical sense as he does). The word heritability has been used for some time by psychologists studying twins, but the measures—such as Holzinger's H-index and various modifications thereof—have not usually corresponded to the geneticist's definition. Jensen has done a great deal to clarify this point.

Heritability, in the geneticist's terms, can be described in three equivalent ways, depending on whether the viewpoint is that of analysis of variance, regression, or correlation analysis: (1) the ratio of the genetic variance to the total variance, (2) the regression of genotype on phenotype, and (3) the square of the correlation of genotype with phenotype. Jensen uses mainly the first.

As he says, the total or phenotypic variance (V_P) can be analyzed into genotypic (V_H) and environmental (V_E) fractions:

$$V_P = V_H + V_E + \text{interactions and error.}$$

The genotypic variance (V_H) can be subdivided further into the additive or genic variance (V_G), dominance (V_D), and inter-locus interaction (epistasis) (V_I). (See Jensen, p. 37.) Thus,

$$V_H = V_G + V_D + V_I.$$

Genetic Theories and Influences
JAMES F. CROW

V_G is defined as the best linear representation of the phenotypic values (best in the least squares sense), and V_D and V_I are treated as deviations from it. This procedure for subdividing V_H has two important advantages: The first is that V_G, V_D, and V_I defined this way are independent and we do not have to worry about covariances among them.

The second advantage is that V_G provides a means for predicting future generations when there is selection. The reason for this lies in the nature of the Mendelian mechanism. What is transmitted by a parent to his progeny is not an intact genotype, but a random sample of genes. Therefore the best prediction is a linear estimate of the average value of the contribution of the individual genes, the variance of which is measured by V_G. I might add, parenthetically, that the situation is not really as tidy as the above sentences may imply. It is difficult to identify the contribution of epistasis, particularly when one considers the complications of linkage between genes on the same chromosome. In many cases the breeder gets satisfactory predictions by simply ignoring epistasis, a fact which may be caused by one of several conditions. It may be that the gene loci act approximately additively on the chosen measurement; it may be that various gene interactions are in opposite directions and therefore cancel each other; or, as is often the case, it may be that the numbers are small, so that even a large discrepancy is regarded as a satisfactory fit, simply because there is not enough statistical power to make a finer discrimination.

The other interactions, between genotype and environment and between environmental components, are not automatically taken care of and have to be considered specifically. It is important conceptually to distinguish, as Jensen does, between *interaction* of heredity and environment (as when a good genotype gets more of a boost from a good environment than a poor genotype does) and *covariation* of heredity and environment (when a good genotype tends to be located where the environment is good). The components due to errors of measurement can usually be ignored if the correlations are corrected for attenuation.

It is important to emphasize that heritability can be defined in two ways:

$$\text{heritability in the narrow sense: } h^2 = \frac{V_G}{V_P}$$

$$\text{and heritability in the broad sense: } H^2 = \frac{V_H}{V_P}.$$

For mnemonic convenience, I shall use h^2 to indicate the narrow definition, which is always quantitatively smaller, and H^2 for the broader (and larger) definition.

Finally, I designate the environmental fraction of the variance as

$$E^2 = \frac{V_E}{V_P}$$

The plant or animal breeder is interested in h^2 because it helps him to predict the expected gains from selection and to estimate the effectiveness of a breeding program. The psychologist is likely to be more interested in H^2 (and E^2) because it partitions the variance into genotypic and environmental components and may thereby afford some insight into each cause. E^2 gives some guidance as to the amount of influence that environmental differences are having and, among these, specific factors may be identified. (See Jensen's discussion of heritability, pp. 33-43.)

Animal and plant experiments have shown that heritability estimates have reasonably good predictive accuracy when the numbers and statistical design are such as to provide a powerful test. However, the prediction is valid for only this particular situation, because heritability is a function of gene frequencies, the mating system, and existing environmental influences. As such, it will change when these change. This means, among other things, that the initial heritability will not be a good guide for long-time selection programs. The program, if successful, will change gene frequencies, and therefore the heritability may change. Furthermore, the environment may change, and this can also change the heritability.

How Valid Are Heritability Measurements of Intelligence?

The animal or plant geneticist gets rid of some of the most troublesome covariances by experimentally designed randomization. This is clearly out of the question in dealing with man. Correlations between relatives are caused by both genetic and environmental similarities. Jensen's general formula for H^2—a great improvement over those of earlier authors, in my opinion—is

$$H^2 = \frac{r_1 - r_2}{\rho_1 - \rho_2} ,$$

where r_1 and ρ_1 are the observed and theoretical correlations for one degree of relationship, and r_2 and ρ_2 are the corresponding quantities for another degree. If we ignore interaction, the correlation will be

$$r_i = \rho_i H^2 + \rho_i' E^2 ,$$

where ρ_i' is the environmental correlation for the ith degree of relationship. As a simple illustration of what happens when the ρ's are not independent, suppose that $\rho_i' = k\rho_i$, where k is a constant. Then

$$\frac{r_1 - r_2}{\rho_1 - \rho_2} = H^2 + kE^2 ,$$

which, instead of measuring H^2, includes an unknown fraction of the environmental variance, thus limiting the usefulness of such a formula. It is likely, for example, that cousins have environmental similarities that are less than those for siblings but more than for children in unrelated families.

If the formula is used to compare monozygous and dizygous twins, there are the often-discussed uncertainties as to whether intra-family environmental differences are the same for the two kinds of twins, as the formula assumes. Identical twins may more often share experiences than dizygotic twins. But, as many authors have pointed out, environmental similarity for monozygotic twins is not necessarily greater than for dizygotic, especially when intra-uterine environment is considered. For example, the likelihood of an unequal blood supply is greater in monozygous twins. Finally, the value of ρ is uncertain. For monozygotic twins it is clearly 1. But for dizygotic twins it is not known exactly. It is roughly 1/2—but decreased by dominance and epistasis and increased by assortative mating, both by unknown amounts.

Most of these difficulties could be removed if individuals of close genetic relationship could be randomized with respect to environments and if unrelated children could be reared in identical environments. These conditions are partially met by studies of twins and siblings reared in different households and by unrelated children reared in the same. As Jensen mentions, the Burt study appears closest to the ideal of placing the separated identical twins into random environments at an early age. There is some reassurance to the skeptic (such as I have been) in that H^2, as estimated by the correlation of one-egg twins reared apart, and E^2, estimated by the correlation between unrelated children reared together, add up to approximately 1—as they should if everything is simple (.75 + .24 = .99 in Jensen's Table 2; .86 + .25 = 1.11 in Burt's individual measurements). Other crosschecks are also in reasonable agreement, but the numbers are small.

If we take the results from many investigations at face value, there is a great deal of consistency, as Jensen points out, and H^2 averages about .8. Furthermore the dominance and epistatic components appear to be small. That the heritability is large is a justifiable conclusion at this stage, although the precise value

77-942 O - 72 - 40

must remain in doubt for the various reasons given. We shall have to be content with measures that are only approximate, pending more evidence on the reality of the assumptions. I agree with Jensen in deploring an uncritical assumption that only environmental factors are important and that genetic differences are negligible.

I admire the diligence of Newman, Shields, and Burt in finding twins and siblings reared apart. Particularly useful, as Jensen has emphasized, would be data on half-siblings reared in different homes. Any excess of similarity of those with a common mother over those with a common father would provide a measure of prenatal and neonatal maternal influences. Though difficult to obtain, it would also be necessary to have data on the correlations between the non-common parents. There will always be some doubt, however, as to whether children from broken homes—separated twins and siblings—and from foster home environments can be regarded as representative of the normal population.

How Important Is It to Measure Heritabilities?

I share Dr. Jensen's interest in trying to determine H^2 and h^2, especially if this information can be extended to other populations. Particularly interesting is his suggestion that heritability be used as one criterion of the culture-fairness of a test. At the same time there are many social decisions that do not depend on a precise knowledge of the heritability of intelligence.

If society decides to improve IQ by eugenic means, h^2 will be useful in providing estimates of the expected gain. I believe that we already know enough to predict that a selection program to increase IQ or *g* would work. There would be an increase, but the amount would be uncertain, because of uncertainties both in the true value of h^2 and in the asumptions underlying its use as a predictor. However, society is clearly not ready to embark on a eugenic program of sufficient scope to make very much difference, even if heritability were equal to 1.

What guidance does H^2 (or E^2) offer for predicting the effect of improvement in the environment? E^2 tells us how much the variance would be reduced if the environment were held constant. It does not directly tell us how much improvement in IQ to expect from a given change in the environment. In particular, it offers no guidance as to the consequences of a new kind of environmental influence. For example, conventional heritability measures for height show a value of nearly 1. Yet, because of unidentified environmental influences, the mean

height in the United States and in Japan has risen by a spectacular amount. Another kind of illustration is provided by the discovery of a cure for a hereditary disease. In such cases, any information on prior heritability may become irrelevant. Furthermore, heritability predictions are less dependable at the tails of the distribution.

A high heritability of intelligence does not necessarily mean that a program of compensatory education is destined to fail, although it may necessitate a larger or more innovative environmental change than if E^2 were larger. Measuring heritability may be less important than getting empirical data on the effects of specific environmental factors. If environment acts as a threshold, as Dr. Jensen suggests, then it would be especially important to identify environmental influences that may be of great influence at the end of the scale, but less so within the normal range.

I am not acquainted with the compensatory education studies Jensen reviews nor am I professionally competent in that area, but my view from the outside is that we should not give up too easily. Perhaps the programs are too little and too late. There are surely a variety of ways of intensifying and improving the effectiveness of education. Also a small change in IQ, especially if accompanied by increased motivation and achievement, may be of great social benefit. Jensen expresses much the same idea:

> Thus it seems likely that if compensatory education programs are to have a beneficial effect on achievement, it will be through this influence on motivation, values, and other environmentally conditioned habits that play an important part in scholastic achievement, rather than through any marked direct influence on intelligence *per se*. The proper evaluation of such programs should therefore be sought in their effects on actual scholastic performance rather than in how much they raise the child's IQ. (p. 59)

Group Differences, Especially Racial Differences

Heritability studies have been confined almost exclusively to white populations and largely to normal environments. How relevant are they to other populations and environments? We are currently especially concerned about culturally disadvantaged groups and racial minorities. Strictly, as Jensen mentions, there is no carryover from within-population studies to between-population conclusions.

I agree that it is foolish to deny the possibility of significant genetic differences between races. Since races are characterized by different gene frequencies, there

is no reason to think that genes for behavioral traits are different in this regard. But this is not to say that the magnitude and direction of genetic racial differences are predictable.

It is clear, I think, that a high heritability of intelligence in the white population would not, even if there were similar evidence in the black population, tell us that the differences between the groups are genetic. No matter how high the heritability (unless it is 1), there is no assurance that a sufficiently great environmental difference does not account for the difference in the two means, especially when one considers that the environmental factors may differ qualitatively in the two groups. So, I think, evidence regarding the importance of heredity in determining group mean differences must come from other kinds of studies.

The failure, thus far, to find identifiable variables that, when matched, will equalize the IQ scores does not prove that the mean difference is hereditary. It can be argued that being white or being black in our society changes one or more aspects of the environment so importantly as to account for the difference. For example, the argument that American Indians score higher than Negroes in IQ tests—despite being lower on certain socio-economic scales—can and will be dismissed on the same grounds: some environmental variable associated with being black is not included in the environmental rating. Behavioral scientists can be expected to disagree, and they do, as to when enough identifiable environmental factors have been shown to be insufficient that the remaining differences should be regarded as mainly genetic. To me, the evidence on this question is not at all conclusive.

Final Comments

One of the goals of a democratic society, I believe, ought to be to provide each individual with the maximum opportunity to satisfy his needs and desires and to contribute to society's betterment through his special abilities. A population with a variety of phenotypes (and genotypes) ought to be more rewarding, and certainly more interesting, than one that is homogeneous. I do not go to the extreme of saying that all variation should be encouraged—I shall be quite happy if the gene for muscular dystrophy becomes extinct—but in general I believe that diversity is good, not bad. In any case, we have it.

Society should recognize that there is a great deal of genetic variability for all kinds of traits, including intelligence and special talents. I think that J. B. S.

Genetic Theories and Influences
JAMES F. CROW

Haldane once said that liberty is the practical recognition of human variability. We should also realize that to whatever extent society is successful in its goals of providing equality of opportunity, to that extent the heritability will increase. In view of this fact, I fully agree with Jensen that, rather than uniformity, the goal should be diversity of educational opportunity with maximum individual opportunity for finding the right niche, and that the reality of individual differences need not and should not mean rewards for some and frustration for others.

The Future of Individual Differences

CARL BERETIER, *The Ontario Institute for Studies in Education*

Professor Bereiter concurs in Jensen's re-emphasis of the heritability of intelligence, but he draws different conclusions about the probable future. Because most intellectual tools which can be learned act as amplifiers rather than equalizers of basic differences in problem-solving ability and because our complex society increasingly emphasizes intelligence rather than other abilities, Bereiter believes that the kind of educational effort recommended by Jensen may in fact only increase the consequences of individual differences. Nevertheless, he suggests that this pessimistic projection may be open to revision in the light of ongoing work in early remedial education.

I have read Dr. Jensen's paper as an essay on the subject of what lies beyond attainment of equal educational opportunity. He does not deny that educational inequities exist and should be rectified, but he is concerned that people seem to expect the removal of such inequities to eliminate effectively the great spread of individual differences in intelligence, with its host of social concomitants. Dr. Jensen expects that in reality the removal of these inequities will have little effect on the spread of individual differences and he proposes that we start developing educational approaches that accept this spread of abilities as hard fact.

My own view of the future of individual differences and their social consequences is even less optimistic than Dr. Jensen's. The heritability of intelligence is unquestionably high, but what is more to the point is that with further social progress its heritability can only increase, because of the elimination of such sources of environmental variance as differences in the quality of education, nutrition, and medical care. One's view of the future beyond equality of opportunity

Harvard Educational Review Vol. 39 No. 2 Spring 1969

The Future of Individual Differences
CARL BEREITER

must, therefore, be of a future in which differences in intelligence are virtually one hundred percent determined by heredity.[1]

The magnitude of these differences then becomes a crucial question; however, it is not magnitude in terms of IQ points that counts, but magnitude in terms of differences in effective problem-solving capability.[2] These are not the same thing, even if they are perfectly correlated. We may expect that through continued scientific progress, through the continued development of intellectual "tools" such as language, logic, thinking machines, and scientific techniques, man's ability to solve problems will increase at an accelerating pace, even though his IQ changes not a jot. There is nothing paradoxical in this. Through the development of mechanical tools man's ability to lift weights, hurl objects, and so on, has been multiplied manyfold, independently of any increase in his basic muscular strength.

Tools, then, may act as amplifiers or equalizers with respect to basic human capabilities. A lever, for instance, amplifies force. If it triples the force Smith and Jones can exert, it also triples the *difference* between the forces Smith can exert and Jones can exert. An electric hoist, on the other hand, is an equalizer. So long as they both have what it takes to operate the hoist, both Smith and Jones can lift the same weight, regardless of their differences in muscular strength. Technology has generally been moving toward tools of the latter type, thus giving rise to the spectre of a future automated world in which individual human differences will no longer count for anything, having been obliterated by the uniformity of machine performance. I would edit this vision in only one important respect: in this future world the overwhelming variable of individual differentiation will be that of intelligence as manifested in the ability to use those tools that make other individual differences irrelevant. This statement rests on the following arguments:

1. The equalizing effect of sophisticated tools is gained by having intelligence take over the function of other abilities. Whereas the photographer once needed the ability to judge depths and levels of illumination accurately in order to take a clear picture, he can now be quite deficient in these abilities providing he is intelligent enough to use his equipment properly. This is not to say that sophisticated tools always require more intelligence to operate than primitive ones;

[1] This eventuality is in no wise to be forestalled by individualized instruction or any more libertarian tactic; on the contrary, such approaches should allow inherited differences to reach full flower, as advertised in the slogan, "enabling each child to realize his fullest potential."

[2] Another way of saying this is that it is surface traits rather than source traits of intellectual ability that count socially (Cattell, 1950). Both, of course, are phenotypes.

they may require less. The point is that the sophisticated tool requires only intellectual ability, whereas the primitive tool required intelligence plus physical strength, manual dexterity, sensory acuity, etc.

2. Intellectual tools, by which I mean algorithms, principles, systems, and devices that are used in processing information, appear in the long run always to function as amplifiers rather than equalizers of intelligence and, thus, to magnify rather than nullify individual differences in ability. The class of tools called mathematics furnishes the clearest examples. Using only arithmetic, people will differ considerably in the complexity of problems they can solve with it, as shown for instance by their performance on the Arithmetic Reasoning Test of the Wechsler Adult Intelligence Scale. Using elementary algebra, the less capable person will be able to accomplish little more than he could with arithmetic, whereas the more capable will be able to use it to solve problems of quite a high level. Using analytical geometry and calculus, the less capable person will again show little gain, except for being able to solve integration and differentiation problems that are clearly set up for him, whereas the more capable person will now be able to solve problems that the duller one cannot even conceive of.

Even when a new tool serves as an equalizer with respect to ability to solve a certain kind of problem, its overall effect seems to be that of an amplifier. Computer programs for the rotation of factors in factor analysis and for the simplification of electrical circuit designs have taken over tasks that used to require considerable art and intelligence. But the ultimate effect of such a development is simply to take out of the hands of specialists and to make available for more general use tools which the intelligent person can use intelligently and the unintelligent person can use unintelligently, thus increasing their manifest difference.

3. Every tool requires certain minimum abilities of a person in order for him to use it at all. Accordingly, each new tool drives a wedge between those who can learn to use it and those who cannot. The more powerful the tool the wider the wedge. An enormous effective gap, for instance, separates those who can learn to read from those who cannot. The social importance of such gaps seems, however, to depend not only on their size, but on where they occur in the distribution of abilities—on whether they separate a small minority at the top or bottom of the distribution or whether they separate the population more nearly into halves.

The direction of progress in the development of intellectual tools and of methods for teaching their use is generally toward lowering the level of intel-

lectual ability prerequisite to their use. Thus, I do not see Dr. Jensen's proposal, that educators look for ways to make school learning less dependent on intelligence, as a very radical one. This is what efforts at curriculum improvement, remediation, and improvement of teaching methods all try to do, whether successfully or not.[3] Bringing intellectual tools within the reach of more people does not necessarily have a leveling effect, however. To replace a method of reading instruction in which fifteen percent of children fail by one in which only five percent fail would tend to make for more social equality; but to take some powerful intellectual tool that could be mastered by only one person in a hundred and to make it so that half the population could master it would be a divisive influence. One such possibility is suggested by the oft-mentioned prospect of household computers. Presumably, such computers would be so simple in their routine operation that all but the most incompetent could manage them. But being general purpose computers, they would lend themselves to all sorts of non-routine applications and thus would provide an intellectual tool of enormous power and versatility to the person who could program them—and this is an ability that might well be put within the reach of about half the population, and yet remain unattainable by the other half. Thus, a sharp and conspicuous split in effective problem-solving ability would arise where none exist at this time.

The magnitude of individual differences referred to in the above arguments is, of course, a subjective matter, a matter of what people make of perceived differences rather than of objective magnitude. There is no absolute sense in which one could say that individual differences in problem-solving ability are greater than, say, individual differences in perceptual abilities; yet as a statement about the recognized and pragmatically significant differences among people in modern societies, it is obviously valid. We are sharply aware of differences among our fellow men in problem-solving ability; such perceived differences figure in countless decisions, with the result that a person's problem-solving ability enters prominently and complexly into the determination of his social fate. Differences in perceptual abilities, on the other hand, come to our attention only rarely and in special circumstances and for most people play little part in determining the course and character of their social lives. It is easy to imagine a world in which the tables would be turned.

[3] Even when new mathematics and science curricula demand more intelligence of the students than old curricula, if one considers what the new curricula are actually trying to teach, it will be seen that they are trying to bring within the reach of a wider population concepts and tools that were previously reserved for a more advanced or gifted minority.

Dr. Jensen has noted that there are societies in which differences in *g* do not count for so much. Presumably, there is a level of description at which it may be said that individual differences in intelligence are of about the same magnitude in those societies as they are in ours, but the level of description is assuredly not that of manifest effectiveness in solving real-life problems or of differentiation of social status on the basis of competence. A common interpretation of this anthropological fact is that, for one or another creditable reason, some other societies do not "value" intellectual abilities as we do. Another interpretation is that these societies lack the intellectual tools that would amplify differences in problem-solving abilities to the point where they are as conspicuous as they are in our society.

Either way, as one moves from relatively primitive to relatively advanced societies, individual differences in intelligence become at once more conspicuous and more consequential in manifold ways. What I have been proposing is simply that this trend will continue into the future at a rapidly accelerating rate, as differences in intelligence take the place of more and more other, formerly compensating, differences in ability and as more and increasingly powerful intellectual tools become available to magnify differences in effective intellectual ability.

We may now hasten to the denouement of this pessimistic story. The prospect is of a meritocratic caste system, based not on arbitrary distinctions of privilege as in traditional caste systems, but on the natural consequences of inherited differences in intellectual potential. These consequences, however, could be expected to extend well beyond differences in occupational status, to include associated differences in attitudes, interests, and ways of life. Assortative mating could be expected to intensify under these conditions, thus leading to further augmentation of inherited differences and rigidification of the caste hierarchy. Such a caste system would be far more resistant to democratizing influences than imposed caste systems of the past. It would tend to persist even though everyone at all levels of the hierarchy considered it a bad thing. The already high level of assortative mating on intelligence, which according to Dr. Jensen is higher than on any other trait that has been investigated, is perhaps the strongest single piece of evidence that progress toward this caste system is already well advanced.[4]

[4] Michael Young's otherwise compelling fictional account of *The Rise of the Meritocracy* (1958) misses the mark, I believe, in focussing on the tyrannical use of IQ tests to fix people's places in the meritocratic hierarchy. Testing is a red herring in this discussion, for it could at best be used only to facilitate discriminations that would be made anyway. The great improvements in intelligence testing that Young saw as necessary to the fullest development of meritocracy are not only unnecessary but also unlikely. The validity of intelligence tests has not increased appreciably in thirty years, and there is little prospect that they will ever account for more than about half the variance in non-test criteria of achievement. Improvements in instructional tech-

The Future of Individual Differences
CARL BEREITER

In this futuristic context, and in light of the failure of education to date, it may seem gratuitous to raise the question of whether education can do anything to equalize effective intelligence (once it has accomplished the still far from realized goal of giving everyone equal access to intellectual tools). I am encouraged to keep the question open, however, if only because of the results of early education experiments that I have had a part in. The approach that I and my co-workers have taken to early education of disadvantaged children has been rather close to that which Dr. Jensen advocates. We were not trying to "stimulate the growth of intelligence," but rather to teach academic skills directly in ways that did not demand of the children abilities they demonstrably did not possess (Bereiter, Engelmann, Osborn, and Reidford, 1966; Bereiter and Engelmann, 1966). As judged by achievement tests, the efforts have been quite successful (Bereiter, 1968), and I think they lend support, at least as far as the early stages of subject-matter learning are concerned, to Jensen's conviction that "all the basic scholastic skills can be learned by children with normal Level I learning ability, provided the instructional techniques do not make *g* ... the *sine qua non* of being able to learn" (p. 117).

Nevertheless, in spite of the fact that the program was never intended to raise IQ and that two-thirds of it was devoted to reading and arithmetic instruction having little or nothing to do with the skills called for on IQ tests, significant IQ gains have been regularly obtained. Over the last four replications they have averaged about 15 points. This seems to be too much of a gain to write off to test-wiseness and things of that sort, especially since the children's IQs were in the middle nineties to begin with and thus rose to substantially above average.

However, we never entertained any illusions that the instruction was improving the children's brains. The most reasonable interpretation had seemed to be that the IQ gains merely reflected the accelerated learning of some kinds of conceptual content sampled by the IQ test (in all cases the Stanford-Binet). This interpretation has received something of a blow, however, from a recent and as yet unpublished study in which we tried out a new curriculum generated by working backward from the Stanford-Binet to create a universe of content for

nology, which would make instruction continuously adaptive to variations in level, rate, and style of learning, are foreseeable, however, and would render IQ testing irrelevant. They would also have the effect of streaming people into different levels of the meritocratic hierarchy without the least hint of coercion. By imagining a tyrannical system, Young imagined one that could be overthrown by the oppressed. The currently high level of assortative mating on intelligence demonstrates how little meritocracy need depend either on IQ-branding or on official control.

which the Stanford-Binet could be considered a content-valid achievement measure. Going at it in this bald-faced manner, we expected to obtain enormous but, of course, psychologically meaningless IQ gains on the Stanford-Binet. As a check on non-specific effects, we also used the WPPSI as a pre- and post-test, without its content's being known during the experiment either to the curriculum writers or to the teachers. Contrary to expectation, the gains on the Stanford-Binet were not large compared to those regularly obtained with the academically-oriented curriculum—about 12 points, and the gains on the WPPSI were exactly the same as those on the Stanford-Binet.

Tracking down what is actually learned in order to account for IQ gains is likely to prove an arduous and perhaps ultimately thankless task. Our unpublished study does not point to any answer but does suggest that there may be more to educationally-induced IQ gains than meets the eye, whereas we, along with Dr. Jensen, had been inclined to assume that there was less.

Here is one possibility: thinking, as even the behaviorists are coming to admit, must surely consist of very long strings of actions or responses, most of which are never directly subjected to corrective feedback. Thus learning to think, to the extent that it occurs, must occur under less than propitious circumstances. This would constitute a situation in which inherited differences in functioning of an otherwise trivial nature could have profound effects. Slight differences in immediate memory, alertness, etc., could spell the difference between learning and not learning some of the cognitive behaviors involved in thinking, or between learning them early and learning them late. Yet with a little help they might be learned—help of a kind that is not regularly provided by the feedback conditions of either normal or school life. Educational programs that produce substantial IQ gains may have inadvertently managed to teach such behaviors. The temporary character of IQ gains doesn't negate this possibility. Most likely such programs, operating blindly in this regard, could do no more than teach early what would be learned later anyway, so that IQs eventually return to their expected levels.[5]

[5] Whether this will be altogether the case with children educated in our program remains to be seen. It was so for the children in the original pilot group. By the end of second grade their IQs had gone back down to their original level. However, the mean IQ for that group had only risen 10 points after treatment. The second wave, on the other hand, showed a 25 point IQ gain in two years of preschool treatment. By the end of first grade their IQs had declined 11 points, but this still left them with a net gain of 14 points and a mean IQ of 110. A randomly equivalent control group given one year of Head Start-like enrichment gained approximately 8 points and remained at the end of first grade with a net gain of 5 points and a mean IQ of 101 (Karnes, 1969).

The Future of Individual Differences
CARL BEREITER

In order to achieve any lasting neutralization of the inherited tendencies leading to lower IQ, it would be necessary to discover cognitive behaviors which duller people will never learn at all and to find ways of teaching them. I will not pretend to specify any such behaviors, even speculatively, but I will suggest a couple of areas where they might lie. One is an area that may be called preliminary information-processing—what you do with incoming information when you don't yet have enough other information to make intelligent use of it. I would suggest that the dull person doesn't do anything with it most of the time, so that he is said to have an attentional deficit (Zeaman and House, 1963), whereas the intelligent person has learned a number of provisional information-processing moves which at least have the effect of preserving the pieces of information in a form so that they can be assembled later (Payne, Krathwohl, and Gordon, 1967). Another is in the construction of solution models for problems (Gagne, 1966), which even some college students seem to do not merely poorly, but not at all (Bloom and Broder, 1950).

Remedial education, along with remedial genetics and remedial biochemistry, might conceivably have some appreciable effect in reducing the spread of individual differences in intelligence. I see no prospect whatever, however, for a reversal of the tendency for intelligence to take over the function of other human abilities. That tendency is intrinsic in the entire progress of science and technology. The domain in which other human abilities are significant becomes increasingly limited to sports and to arts where the scope of intelligence is arbitrarily restricted (through restrictions on the kind of equipment that may be used, for instance). Thus Dr. Jensen's closing appeal for diversity of aims in education inspires more nostalgia than hope, recalling the nearly vanished era when blacksmith, watchmaker, woodcarver, gardener, and a host of others could attain some measure of distinction on the basis of special abilities little related to general intelligence. Special abilities will continue to have a place, of course, but as adjuncts rather than alternatives to general intelligence.[6] If we are to

[6] Throughout this discussion I have followed Dr. Jensen in using the terms *g*, intelligence, and IQ interchangeably. I don't believe that either his argument or mine would be materially altered by dropping the notion of *g* and adopting a multifactorial view of intellectual abilities, as in Guilford (1967). The main difficulty would be the shortage of relevant data on separate intellectual abilities, compared to what is available on general intelligence. A special drawback to approaching the problem multifactorially is the lack of data on hereditary and environmental contributions to the *correlations* between mental abilities. According to Thompson (1966), this matter has never been studied, although the methodology is available and has been applied to other problems. If we regard *g* as an unrotated first factor (Rimoldi, 1951), its composition would naturally change with change in the selection of tests, as Dr. Jensen notes, and preferred

make something of the "untapped reservoir" of learning ability that Dr. Jensen finds among the disadvantaged, it would seem that we must look—as educators and psychologists have really only just begun to do—for ways to marshal this learning ability to the task of learning to think.

---

selections of tests might change as cultural conditions changed. For instance, there could be a shift toward greater emphasis on creativity measures. Such shifts would have implications for who ranks where in a meritocratic hierarchy but not, foreseeably, of such a radical kind as to require serious qualifications in any arguments presented here.

References

Bereiter, C. A non-psychological approach to compensatory education. In M. Deutsch, I. Katz, and A. R. Jensen (Eds.), *Social class, race, and psychological development.* New York: Holt, Rinehart & Winston, 1968.

Bereiter, C. and Engelmann, S. *Teaching disadvantaged children in the preschool.* Englewood Ciffs, N. J.: Prentice-Hall, 1966.

Bereiter, C., Engelmann, S., Osborn, J., and Reidford, P. An academically-oriented pre-school for disadvantaged children. In F. M. Hechinger (Ed.), *Pre-school education today.* Garden Hills, New York: Doubleday, 1966, pp. 105-135.

Bloom, B. S. and Broder, L. J. *Problem-solving processes of college students.* Chicago: University of Chicago Press, 1950.

Cattell, R. B. *Personality: A systematical theoretical and factual study.* New York: McGraw-Hill, 1950.

Gagne, R. M. Human problem solving: internal and external events. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory.* New York: Wiley, 1966.

Guilford, J. P. *The nature of human intelligence.* New York: McGraw-Hill, 1967.

Karnes, M. B., Teska, J. A., Hodgins, A. S. A longitudinal study of disadvantaged children who participated in three different preschool programs: Traditional, direct verbal, and amelioration of learning deficits. Paper read at American Education Research Association, Los Angeles, February, 1969.

Payne, D. A., Krathwohl, D. R., and Gordon, J. The effect of sequence on programmed instruction. *American Educational Research Journal,* 1967, 4, pp. 125-132.

Rimoldi, H. J. A. The central intellective factor. *Psychometrika,* 1951, 16, pp. 75-102.

Thompson, W. R., Multivariate experiment in behavior genetics. In R. B. Cattell (Ed.), *Handbook of Multivariate Experimental Psychology.* Chicago: Rand-McNally, 1966, pp. 711-731.

Young, M. *The rise of the meritocracy.* London: Thames & Hudson, 1958.

Zeaman, D., and House, B. J. The role of attention in retardate discrimination learning. In N. Ellis (Ed.), *Handbook on Mental Deficiency.* New York: McGraw-Hill, 1963.

Piagetian and Psychometric Conceptions of Intelligence

DAVID ELKIND, *University of Rochester*

Professor Elkind devotes much of his discussion to the concept of intelligence. He finds both similarities and differences when comparing the Piagetian description of intelligence with Jensen's (and the psychometrician's) definition of intelligence. Operating from quite different assumptions than those of J. McV. Hunt (Piaget's Structuralism, rather than neurology) Elkind also finds reason to believe that intelligence is developed in experience. For Piaget and Elkind, intelligence is "an extension of biological adaptation" and is characterized by ability to assimilate (develop in response to internal processes) and accommodate (respond to environmental intrusions).

I have been asked to respond to Professor Jensen's paper from the standpoint of Piaget's genetic psychology of intelligence. While I clearly cannot speak for Piaget, only the "Patron" can do that, I can react as someone steeped in Piagetian theory and research and as one who looks at cognitive problems from the Genevan perspective. Accordingly, while I hope that what I have to say would be acceptable to Piaget, I cannot guarantee that this is in fact the case, and must take full responsibility for whatever is said below. I plan to discuss, in the first section of the paper, some of the similarities between the Piagetian and psychometric positions. Then, in the second section, some of their differences will be pointed out. Finally, in the third section, I want to consider two related practical issues regarding the modification of intelligence.

Harvard Educational Review Vol. 39 No. 2 Spring 1969

Conceptual Similarities

What struck me in reading Professor Jensen's paper, and what had not really occurred to me before, were the many parallels and affinities between the psychometric or mental test approach to the problem of intelligence and the developmental approach as represented by Piaget. It brought to mind the fact that Piaget began his career as a developmental psychologist by working in Binet's laboratory where he sought to standardize some of Burt's (1962) reasoning tests on Parisian children. Indeed, Piaget's *method clinique* is a combination of mental test and clinical interview procedures which consists in the use of a standardized situation as a starting point for a flexible interrogation. The affinities, however, between the Piagetian and psychometric approaches to intelligence run more deeply than that. In this section I want to discuss such affinities: the acceptance of genetic and maturational determination in intelligence, the use of nonexperimental methodologies and the conception of intelligence as being essentially rational.

Genetic Determination

Implicit and often explicit in both the psychometric and Piagetian positions is the assumption that mental ability is, in part at least, genetically determined. With respect to the psychometric position, it assumes that at least some of the variance in intelligence test performance is attributable to variance in genetic endowment (Burt & Howard, 1957, Jensen). Piaget (1967a) also acknowledges the importance of genetic factors for intellectual ability but qualifies this by pointing out that what may be genetic in one generation may not always have been so and could be the partial result of prior environmental influences. So, for Piaget, as for the biologist Waddington (1962a) there is a certain relativity with respect to what is attributed to genetic endowment because what is genetic now may not always have been genetic. To illustrate, Waddington (1962a) observed that after several generations a strain of the fly grub drosophilia developed enlarged anal papillae when reared on a high salt diet. When the insects were returned to a "normal" low salt diet the anal papillae of successive generations became less large but never returned to their original size. Waddington speaks of this as "genetic assimilation" by which he means that the effects of an altered environment upon the selection process within a species may not be completely reversible even when the environment returned to its unaltered state.

One consequence of their joint acceptance of the partial genetic determination of intellectual ability, is that both psychometricians and Piaget recognize the im-

portance of maturation in human development. To illustrate their commonality in this regard, consider these two passages, one written by Harold Jones in 1954 and the other by Piaget in 1967.

Dubnoff's work, together with other related studies, may lead to the speculative suggestion that between natio-racial groups, as within a given group, a slight tendency exists for early precocity to be associated with a slower mental growth at later ages and perhaps with a lower average intelligence level at maturity. A parallel situation may be noted when we compare different animal species; among the primates, for example, the maturity of performance at a given age in infancy can be used inversely to predict the general level of adaptive ability that will be attained at the end of the growth span. (Jones, 1954, p. 638)

And Piaget writes:

We know that it takes 9 to 12 months before babies develop the notion that an object is still there when a screen is placed in front of it. Now kittens go through the same substages but they do it in three months—so they are six months ahead of the babies. Is this an advantage or isn't it? We can certainly see our answer in one sense. The kitten is not going to go much further. The child has taken longer, but he is capable of going further so it seems to me the nine months were not for nothing. (Piaget, 1967b)

Non-Experimental Methodology

In addition to their shared genetic or maturational emphasis, the Piagetian and psychometric approaches to intelligence have still another characteristic in common. This common feature is their failure, for the most part, to use the experimental method in the strict sense of that term. It seems fair to say that most of the studies which attempt to get at the determinants of test intelligence are correlational in nature. By and large such studies attempt to relate the test scores of parents and their children, of twins or of adopted children and their parents, or of the same children tested at different points in time and so on. Only in rare instances such as the Skeels (1966) study is an attempt made to modify intelligence by active intervention and with the utilization of a control group which does not receive the experimental treatment. While experimental work on human intelligence might well be desirable, such research often raises serious moral and ethical questions.

Piaget, for his part, has not employed the experimental method simply because it was not appropriate for the problems he wished to study. This is true because Piaget has been primarily concerned with the diagnosis of mental contents and abilities and not with their modification. To illustrate, the discovery of

77-942 O - 72 - 41

what the child means by "more," "less" and "same" number of things requires flexible diagnostic interview procedures and not experimental procedures. Once the concept is diagnosed, then experimental methods are appropriate to determine the effects of various factors on the attainment and modification of the concepts in question. The sequence of events is not unlike the situation in medicine where the discovery or diagnosis of a disease is often the first step to its experimental investigation. In short, Piaget has focused upon the discovery of what and how children think and not with the modification of thinking which is a subsequent and experimental question. In every science there is a natural history stage of enquiry during which relevant phenomena must be carefully observed and classified. American psychology has often tried to bypass this stage in its headlong rush to become an experimental science. In his studies Piaget has revealed a wide range of hitherto unknown and unsuspected facts about children's thinking, which have in America now become the starting points for a great deal of experimental investigation. What is often forgotten, when Piaget is criticized for not using the experimental method, is that such a method would not have revealed the wealth of phenomena which experimental investigators are now so busily studying.

Rationality as the Definition of Intelligence

There is a third and final commonality in the mental test and Piagetian approaches to intelligence which should be mentioned. This commonality resides in what these two positions regard as the nature or essence of intelligence. While there is considerable variability among psychometricians in this regard, many agree in general with the position taken by Jensen (1969). Jensen argues that the *g* factor which is present in all tests of mental ability appears in its purest forms on tests of generalization and abstraction. Spearman (1923) called these activities the education of relations (A is greater than B; B is greater than C; so A is in what relation to C?) and of correlates (Complete the series A AB ABC ———). While intelligence tests contain measures of many different types of mental abilities, including language and perceptual skills, the psychometric approach holds that the most central feature of human intelligence is its rationality, or as Wechsler put it: "Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Wechsler, 1944, p. 3).

For Piaget, too, the essence of intelligence lies in the individual's reasoning capacities. Piaget, however, is more specific in his description of these abilities

and defines them in terms of mental operations which have the properties of mathematical groupings in general and the property of reversibility in particular. An operational grouping is present when in the course of any mental activity one can always get back to the starting point. For example, if the class *boys* and the class *girls* is mentally combined to form the class *children*, it is always possible to recapture the subclass by subtraction. That is to say, the class of children minus the class of boys equals the class of girls. Put differently, the operation of subtraction can be used to undo the operation of addition so that each of the combined classes can be retrieved. Verbal material learned by heart is, however, not rationally organized as is illustrated by the fact that no matter how well a passage is learned, it is impossible, without additional effort, to say it backwards. If an operational system were involved, having learned the passage forward would automatically imply the ability to say it backwards. In Piaget's view, neither perception nor language are truly rational since neither one shows complete reversibility. So, while perception and language play an important part in intellectual activity, they do not epitomize that activity.

The psychometric and Piagetian approaches to intelligence thus agree on its genetic determination (at least in part), and on the use of non-experimental methodology and upon the essentially rational nature of mental ability. After this look at their commonalities, it is perhaps time to look at their differences.

Conceptual Differences

Despite the commonalities noted above, the psychometric and developmental approaches to intelligence also differ in certain respects. These differences, however, derive from the unique ways in which the psychometricians and Paiget approach and view intelligence and not from any fundamental disagreements regarding the nature of intelligence itself. In other words the differences are due to the fact that the two approaches are interested in assessing and describing different facets of intelligent behavior. Accordingly the differences arise with respect to: (a) the type of genetic causality they presuppose; (b) the description of mental growth they provide; and (c) the contributions of nature and nurture which they assess.

Genetic Causality

Although the Piagetian and psychometric approaches to intelligence agree on the importance of genetic determination, at least in part, of human mental ability, each approach emphasizes a somewhat different mode of genetic deter-

mination or causality. In order to make these differences clear, it is necessary to recall some of the basic features of evolutionary theory upon which all modern conceptions of intelligence are based.

Within the Darwinian conception of evolution there are two major phenomena that have to be taken into account: within-species variability and natural selection. For any given species of animal or plant one can observe a range of variations in such features as color, shape and size. Among a flock of robins, to illustrate, one can see that some adult birds differ in size, in richness of breast coloration and that some even manifest slight variations in head and wing conformation. Similar variations can be observed among a group of collies, Persian cats and even among tomato plants in the garden. This within-species variability, we know today, is due to the chance pairings of parental genes and to gene complexes which occur because each parent contributes only half of its genetic complement to its offspring. Variations within a given species at a given time are, therefore, primarily due to chance factors: namely the random genetic assortments provided by the parent generation. One determinant of variability among animals and plants is then, simply, chance.

Now in the psychometric conception of intelligence, this random type of variation is just what is presupposed. Test intelligence, it is assumed, is randomly distributed in a given population at a given time and such distributions should resemble the bell shaped curve of the normal probability function. Measurement of human abilities does in fact reveal a tendency for such measurements to fall into normal distributions. In addition evidence such as "regression toward the mean" (children of exceptionally bright or dull parents tend to be less bright and less dull than their parents) is also characteristic of genetic traits which are randomly determined. In short, when the psychometrician speaks of genetic determination, he is speaking of the chance gene combinations which produce a "normal" bell-shaped distribution of abilities within a given population.

Obviously this description of genetic determination is extremely over-simplified; we know that a test score is a phenotype which is determined by many different factors not all of which are genetic. Jensen, to illustrate, breaks down the variance of test intelligence into a large number of components such as genotypic variation, environment, environment genotype interaction, epistasis, error of measurement variance and so on. With the exception, perhaps, of the selective mating variable, however, all of these factors can again be assumed to operate in a random manner so that one might say that the chance distribution of observed test scores is the product of many underlying chance distributions. That the psy-

chometric approach does in general presuppose a random distribution is also shown by the fact that the criterion of a true change in intellectual ability is the demonstration that such a change could *not* be attributed to chance factors.

That variability within a species is in part determined by chance gene and gene complex assortments has of course been demonstrated by Mendel and all of the research which has derived from his theory of genetics. There are, however, other forms of organismic variability which cannot be attributed to chance. Natural selection, the other component of evolution, is never random but always moves in the direction of improved adaptation to the milieu. To illustrate, over the past hundred years there has been a gradual predominance of dark over light colored moths in the industrial sections of England. Kettlewell (1955) demonstrated the survival value of dark coloration by showing that light moths placed on soot darkened bark were more readily eaten by insectivorous birds than were similarly placed dark moths. When variations across generations are considered, the variations are not random but rather show a clear cut direction.

The same holds true within the course of individual development. In the case of individual growth, however, the direction of progress is not determined by mating practices but rather by biochemical mechanisms which are only now in the process of being understood. That these biochemical agents determine the direction of development, however, cannot be doubted. As Waddington (1962b) points out, animals consist of a limited variety of cells such as nerve cells, muscle cells and so on. Likewise the organs of the body are also distinct from one another in form, composition and function. What direction particular cells will take as the egg matures will depend upon the action of chemical agents which Spemann (discussed in Bertalaffny, 1962) called *organizers* with definite loci in the cell material called *organization centers.* It is the organizer which determines whether particular cells will become nerve, muscle or organ tissue. Individual development, therefore, is not determined by random factors but rather by biochemical organizers which specify the nature and direction of organismic differentiation.

Now when Piaget speaks of the genetic determination of intelligence, he has in mind not the random factors which determine gene combinations, but rather the non-random action of biochemical organizers and organization centers. Indeed, this is the kind of determination which Piaget assumes when he argues that the *sequence* in which the child attains the successive components of a concept or in which he acquires systems of mental operations, is invariant. In the formation of body organs the order of differentiation is fixed because each new

phase of differentiation produces the organizer for the next stage. In Piaget's view this is equally valid for the growth of cognitive structures because the preceding cognitive structures, say the concrete operations of childhood, are a necessary prerequisite to the elaboration of the more complex formal operational structures of adolescence. For Piaget, then, genetic determination means that there are factors which give development a definite non-random direction.

In pointing out that the Piagetian and psychometric approaches to intelligence postulate different forms of genetic determinism, I want to reiterate that these two positions are not in contradiction one with the other. The mental test approach to intelligence is concerned with inter-individual differences in ability and these are, in so far as we know, largely randomly determined. Piaget, in contrast, is concerned with the intra-individual changes which occur in the course of development and these, to the best of our knowledge, are not random but rather have a direction given them by specific organizing mechanisms. Accordingly, and this is the genius of evolution, human intelligence manifests both determinism *and* freedom.

The Course of Mental Growth

Let us look now at a somewhat different issue, the age-wise course of mental growth. Here again we find a difference in perspective rather than a contradiction in conception as between the two positions. In psychometric terms, the course of mental growth is plotted as a curve which measures the amount of intelligence at some criterion age that can be predicted at any preceding age. As Bloom (1964) has pointed out, when age 17 is taken as the criterion age, some 50% of the total IQ at that age can be predicted at age four, and an additional 30% can be predicted from ages four to eight. Based on correlational data of this sort, curves of mental growth appear to rise rapidly in early childhood and taper off to a plateau in late adolescence. Such curves, it must be noted to avoid a frequent misinterpretation, say nothing as to the *amount* or *quality* of knowledge at given age levels. (See Jensen, pp. 115-117.)

From the mental test perspective, therefore, intellectual growth is pretty much a statistical concept derived from correlations of test scores obtained at different age levels on the same individuals in the course of longitudinal studies. Such curves can be interpreted as reflecting the rate of mental growth but say nothing as to the nature of what is developing. Indeed, if intelligence is defined in the narrow sense of the abilities to generalize and abstract, then any qualitative differences in these abilities will necessarily be obscured by the curve of mental

growth which suggests merely a quantitative increase in mental ability with increasing age.

Looked at from the standpoint of Piagetian psychology, however, mental growth involves the formation of new mental structures and consequently the emergence of new mental abilities. The child, to illustrate, cannot deal with propositional logic of the following sort, "Helen is shorter than Alice and taller than Ethel, who is the tallest of the three?" (Glick & Wapner, 1968), nor can children grasp the metaphorical connotations of satirical cartoons or proverbs (Shaffer, 1930). Adolescents, in contrast, have no trouble with either propositional logic or with metaphor. In the Piagetian view, therefore, mental growth is not a quantitative but rather a qualitative affair and presupposes significant differences between the thinking of children and adolescents as well as between preschool and school age children.

These qualitative differences are, as a matter of fact, built into the items of mental tests but are masked by the assignment of point scores to successes and failures. On the Wechsler Intelligence Scale for Children various of the subtests recognize qualitatively different responses only by assigning them additional points (Wechsler, 1949). For example, a child who says that a peach and a plum are alike because "they both have pits" is given a single point, whereas a child who says "they are both fruit" is given two points. On other sub-tests, such as the arithmetic sub-test, there is no point differential for success on problems which patently require different levels of mental ability. To illustrate, correct answers to the following two problems are both given only a single point: "If I cut an apple in half, how many pieces will I have?" A correct answer to that question is given the same score as the correct answer to this problem:

> Smith and Brown start a card game with $27 each. They agree that at the end of each deal the loser shall pay the winner one third of what he (the loser) then has in his possession. Smith wins the first three deals. How much does Brown have at the beginning of the fourth deal?

Clearly, the items on any given sub-test can tap quite different mental processes but these qualitative differences are obscured by assigning equivalent point scores to the various items regardless of the mental processes involved.

This is not to say that Piaget is right and that the mental test approach is wrong, or vice versa. The quantitative evaluation of mental growth is necessary and has considerable practical value in predicting school success. The qualitative approach is also of value, particularly when diagnosis of learning difficulties and

educational remediation are in question. Which approach to mental growth one adopts will depend upon the purposes of the investigation. The only danger in the quantitative approach is to assume that, because sub-tests include items of the same general type and are scored with equal numerical weights, that they therefore assess only quantitative differences in the ability in question.

The Contributions of Nature and Nurture to Intelligence

Still a third way in which the psychometric and Piagetian views of intelligence differ has to do with the manner in which they treat the contributions of nature and nurture to intellectual ability. In the psychometric approach this contribution is treated substantively, with regard to the amount of variance in intellectual ability that can be attributed to nature and nurture respectively. Piaget, on the contrary, treats these contributions functionally with respect to the regulative role played by the environment or inner forces for any given mental activity. Both positions now need to be described in somewhat more detail.

The psychometric approach is substantive (and static) in the sense that it regards intelligence as capable of being measured and holds that such measures can be used to assess the extent to thich nature and nurture contribute to intellectual ability. In the discussion of genetic causality the various components into which test scores could be analyzed were briefly noted. We are indebted to writers such as Burt & Howard (1957) and Jensen for making clear the many and complex determinants into which test performance can be analyzed. Without wishing to minimize these other determinants, the needs of the present discussion will be served if we consider only how the psychometric approach arrives at the contribution of the heredity and environmental factors.

As Jensen points out, heritability is the proportion of variability among observed or phenotypic intelligence (test scores) that can be attributed to genotypic variations. Estimates of heritability are obtained from correlational data for subjects with known kinship relations such as parents and children, siblings, and identical twins. The contribution of the environment is arrived at somewhat differently. Variability in intelligence test scores attributable to the environment is estimated from that variability which cannot be attributed to any other factors. It is, in fact, the residual variance, that which is left after all the other factors contributing to intelligence test performance have been accounted for. For the psychometrician, then, nature and nurture are regarded as substantive and static, and their contributions are assessed quantitatively with the aid of statistical procedures.

When we turn to the work of Piaget, however, we encounter quite a different conception of the contributions of nature and nurture. In Piaget's view, these contributions must be conceived functionally and dynamically with respect to their regulatory control over various mental activities. In this regard Piaget's views are not unlike those of David Rapaport (1958) who spoke of "the relative autonomy of the ego," a conception which may help to introduce Piaget's somewhat more difficult formulation. Rapaport argued that we are endowed with some mental processes, such as perception, that are responsive to the environment and so tend to guarantee or insure a certain independence of the mind from the domination of instinctual drives. Other mental processes, such as fantasy, are most responsive to internal forces and these in turn guarantee a certain independence of the mind from the domination of the environment. The presence and activity of both types of processes thus insures that the mind is enslaved neither by the environment nor by drives but retains a "relative autonomy" from both.

Piaget's view (1967c) is roughly similar. He argues that intelligence is an extension of biological adaptation which, in lieu of the instinctive adaptations in animals, permits relatively autonomous adaptations which bear the stamp not only of our genetic endowment, but also of our physical and social experience. On the plane of intelligence we inherit the processes of assimilation (processes responsive to inner promptings) and of accommodation (processes responsive to environmental intrusions). Assimilative processes guarantee that intelligence will not be limited to passively copying reality, while accommodative processes insure that intelligence will not construct representations of reality which have no correspondence with the real world. To make this functional conception of the contributions of nature and nurture to intelligence concrete, let us consider several different mental abilities which are differently regulated by internal and external forces.

If we look at imitation (Piaget, 1951), it is clear that it is largely accommodative in the sense that it is most responsive to environmental influence and is relatively independent of inner forces. The vocal mimic, for example, is expert to the extent that he can capture the pitch, timbre and inflections of his model's voice and to the extent to which he can suppress those aspects of his own speech which differ from the model's. Play, in contrast, is largely assimilative in that it is most responsive to inner needs and is relatively independent of environmental influence. The child who uses a stick alternatively as a gun, as an airplane and as a boat has responded to the object largely in terms of his own inner needs and with a relative disregard of its real properties.

Between the two extremes of imitation and play is intelligence which manifests a balance or equilibrium between assimilative and accommodative activities and is thus relatively autonomous both of inner *and* outer forces. To illustrate, suppose we deduce, from the premise that Helen is taller than Jane and that Jane is taller than Mary, that Helen is the taller of three girls. We have in so doing attained a new bit of knowledge, an adaptation, but without altering the elements involved (assimilation without transformation of the objects) and without modifying the reasoning processes (accommodation without alteration of mental structures). Reason, or intelligence, is thus the only system of mental processes which guarantees that the mind and the environment will each retain its integrity in the course of their interaction.

Accordingly, for Piaget as for Rapaport, the question is not how much nature and nurture contribute to mental ability, but rather the *extent to which various mental processes are relatively autonomous from environmental and instinctual influence*. Such a conception is functional and dynamic, rather than substantive and static, because it deals with the regulatory activity of nature and nurture upon various mental processes. Those processes which show the greatest independence from environmental *and* internal regulation, the rational processes, are the most advanced of all human abilities. It is for this reason that Piaget reserves for them, and for them alone, the term intelligence.

In summary then, the psychometric and Piagetian approaches to intelligence differ with respect to: (a) the type of genetic causality which they presuppose; (b) their conceptions of the course of mental growth; and finally (c) the manner in which they conceive the contributions of nature and nurture to intellectual ability. In closing this section on the differences between the two positions I want to say again that the differences arise from differences in perspective and emphasis and are not contradictory but rather complementary. Both the psychometric and the Piagetian approaches to the conceptualization of human intelligence provide useful starting points for the assessment and interpretation of human mental abilities. Let us turn now to a couple of practical issues related to the modification and stimulation of mental abilities.

Practical Issues

In his essay, Jensen has tried to clarify many of the ambiguities regarding the nature and modification of intellectual ability and to put down some of the myths and misinterpretations prevalent with regard to test intelligence. For the most

part, I find myself in agreement with Jensen and in this section, I would like to discuss two practical issues related to the modification and stimulation of intellectual abilities which seem to involve some misinterpretation of the Piagetian position. First, Piaget's insistence upon the qualitative differences between the modes of thinking at different age levels has been wrongly taken to suggest the need for preschool instruction in order to move children into concrete operational stage more quickly. Secondly, Piaget's emphasis upon the non-chance or self-directed nature of mental development has mistakenly been taken as justification for the use of methods such as "discovery learning" which supposedly stimulate the child's intrinsic motivations to learn. I would like, therefore, to try in the following section to clarify what seems to me to be the implications of Piaget's conception of intelligence for preschool instruction and for the implementation of intrinsic motivation.

Preschool Instruction

There appears to be increasing pressure these days in both the popular and professional literature for beginning academic instruction in early childhood, i.e., from 3 to 5 years. Bruner's famous statement that "We begin with the hypothesis that any subject can be taught effectively in some intellectually honest form to any child at any stage of development" (Bruner, 1962, p. 33) as well as the work of Hunt (1961), of Bloom (1964), of O. K. Moore (1961), of Fowler (1968), and of Skeels (1966) have all been used in the advocacy of preschool instruction. Indeed Piaget and Montessori have been invoked in this connection as well. The argument essentially is that the preschool period is critical for intellectual growth and that if we leave this period devoted to fun and games, we are lowering the individual's ultimate level of intellectual attainment. Parental anxiety and pressure in this regard have been so aroused that legislation has been passed or is pending for the provision of free preschool education for all parents who wish it for their children in states such as New York, Massachusetts and California.

What is the evidence that preschool instruction has lasting effects upon mental growth and development? The answer is, in brief, that there is none. To prove the point one needs longitudinal data on adults who did not have preschool instruction but who were equal in every other regard to children receiving such instruction. With the exception of the Montessori schools, however, the preschool instruction programs have not been in existence long enough to provide any evidence on the lastingness of their effects. Indeed, most of the earlier work on the effects of nursery school education (see Goodenough, 1940, and Jones, 1954,

for reviews of this literature) has shown that significant positive effects are hard to demonstrate when adequate experimental controls are employed. It is interesting that no one, to my knowledge, has done a longitudinal study of adult Montessori graduates. Have they done better in life than children from comparable backgrounds not so trained? In any case, it is such unavailable longitudinal data that is crucial to the proposition that the preschool period is a critical one for intellectual development.

I am sure that someone will object at this point that studies of mental growth such as those of Bloom (1964) suggest that half of the individual's intellectual potential is realized by age four. Does this not mean that the preschool period is important for intellectual growth and that interventions during this period will have lasting effects? Not necessarily, if we look at the facts in a somewhat different way. Bloom writes, "Both types of data suggest that in terms of intelligence measured at age 17, about 50% of the development takes place between conception and age 4, about 30% between ages 4 and 8, and about 20% between 8 and seventeen" (Bloom, 1964, p. 88). Now an equally feasible implication of this statement is quite in contradiction to that of preschool instruction: the child has only 50% of his intellectual ability at age 4 but 80% at age 8, why not delay his education three years so that he can more fully profit from instruction? With 80% of his ability he is likely to learn more quickly and efficiently and is not as likely to learn in ways that he will need to unlearn later. That is to say, without stretching the fact, it is possible to interpret the Bloom statement as implying that instruction should *not* be introduced into the preschool program.

Not only is there no clear-cut longitudinal data to support the claims of the lastingness of preschool instruction, there is evidence in the opposite direction. The work cited by Jones (1954) and by Piaget (1967b) in the quotations given earlier in this paper are cases in point. This evidence, together with more recent data reported in Jensen's paper, suggest a negative correlation between early physical maturation and later intellectual attainments. Animals are capable of achieving early some skills (a dog or a chimp will be housebroken before a child is toilet trained) but perhaps at the expense of not being able to attain other skills at all. This data suggests the hypothesis that *the longer we delay formal instruction, up to certain limits, the greater the period of plasticity and the higher the ultimate level of achievement.* There is at least as much evidence and theory in support of this hypothesis as there is in favor of the early-instruction proposition. Certainly, from the Piagetian perspective, there are "optimal periods" for the growth of particular mental structures which cannot be rushed.

Conceptions of Intelligence

DAVID ELKIND

Please understand, I am not arguing against the benefits of preschool enrichment for children. Even preschool instruction may be of value for those disadvantaged children who do not benefit from what Strodtbeck (1967) called the "hidden curriculum of the middle-class home." What I am arguing is that there is no evidence for the *long term effects* of either preschool instruction or enrichment. Nursery school experience most assuredly has immediate value for the child to the extent that it helps him to appreciate and enjoy his immediate world to the full and to better prepare him for future social and intellectual activities. Everyone, for example, recognizes the value of a vacation without expecting that it will produce any permanent alterations. Isn't it enough that we lighten the burdens of childhood for even a brief period each day without demanding at the same time that we produce permanent results? The contributions of the nursery school, no less than that of the vacation, do not have to be long-lived to be of value.

In closing the discussion, I would like to emphasize another side to this issue of preschool instruction. This is the consideration that the emphasis on preschool education has obscured the fact that it is the elementary school years which are crucial to later academic achievement. It is during these years that the child learns the basic tool subjects, acquires his conception of himself as a student and develops his attitudes towards formal education. In this connection it might be well to quote a less publicized finding of Bloom's (1964) study:

> We may conclude from our results on general achievement, reading comprehension and vocabulary development, that by age 9 (grade 3) at least 50% of the general achievement pattern at age 18 (grade 12) has been developed whereas at least 75% of the pattern has been developed by age 13 (grade 7). (Bloom, 1964, p. 105)

With respect to the intellectual operations of concern to Piaget, similar trends appear to hold true. While children all over the world and across wide ranges of cultural and socioeconomic conditions appear to attain concrete operations at about the age of 6 or 7 (Goodnow, 1969), the attainment and use of formal operations in adolescence, in contrast, appear to be much more subject to socioculturally determined factors such as sex roles and symbolic proficiency (Elkind, 1961; Elkind, Barocas & Rosenthal, 1968; Goodnow & Bethon, 1966). Apparently, therefore, environmental variation during the elementary school period is more significant for later intellectual attainments of the Piagetian variety. In short, there is not much justification for making the preschool the scapegoat for our failures in elementary education. Like it or not, the years from six to twelve are still the crucial ones with respect to later academic achievement.

Motivation and Intellectual Growth

In recent years there has been an increasing recognition among psychologists such as Berlyne (1965), Hunt (1965), and White (1959), that certain mental activities can be self-rewarding and do not have to be externally reinforced. European writers such as Piaget (1954) and Montessori (1964) long ago recognized the existence of "intrinsic motivation" (to use Hunt's apt phrase), and Montessori in particular gave incomparable descriptions of children who suddenly discover they can read and proceed to read everything in sight. Piaget (1967d) too, has argued that needs and interests are simply another aspect of all cognitive activities.

Educators, however, in their efforts to capitalize upon this intrinsic motivation seem to have missed the point of what Montessori and Piaget had in mind. To maximize intrinsic motivation and to accelerate mental growth we have recently had an emphasis upon "learning by discovery" and upon "interesting reading materials" and so on. These approaches miss the point because they assume that intrinsic motivation can be built into materials and procedures which will in turn maximize mental growth. But as Piaget and Montessori pointed out (Elkind, 1967) intrinsic motivation resides in the child and not in methods and procedures. It is the child who must, at any given point in time, choose the method of learning and the materials that are reinforcing *to him*. Without the opportunity for student choice and the provision of large blocks of time in which the child can totally engross himself in an activity, the values of intrinsic motivation will not be realized.

Indeed, I am very much afraid that by the time most children have reached the third or fourth grade a good deal of their intrinsic motivation for learning has been stifled. This is because spontaneous interest follows only the timetable of the child's own growth schedule. We can all remember, I am sure, those periods when we were so totally immersed in an activity that we forgot time, food and rest. During such periods we are at our creative and productive best and afterwards the feeling of exhaustion is coupled with a deep sense of accomplishment. In the school, however, we do not permit children to become totally engrossed in an activity but rather shuttle them from activity to activity on the hour or half hour. The result is what might be called *intellectually burned children*. Just as the burned child shuns the fire so the intellectually burned child shies away from total intellectual involvement.

How is this condition produced? In clinical practice we often see children (and adults) who are unwilling to form any emotional attachment. In the history of such children one always finds a series of broken relationships due to a wide vari-

ety of causes including the death of parents or the forced separation from them. Such children have learned that every time they reached out and became emotionally involved, rejection, hurt and misery were the result. Consequently they prefer not to get involved any more because the pain and anguish of still another broken relationship is just too high a price to pay for an emotional attachment. The intellectually burned child is in somewhat the same position. He refuses to become totally involved in intellectual activities because the repeated frustration of being interrupted in the middle is just too much to bear. Our lockstep curricula, thirty minutes for this and an hour for that, have the consequence, I suspect, of producing children who shun the fire of intense mental involvement.

Accordingly, the educational practice which would best foster intrinsically motivated children in the Piagetian and Montessori sense would be the provision of "interest areas" where children could go on their own and for long periods of time. Only when the child can choose an activity and persist at it until he is satiated can we speak of true intrinsically motivated behavior. Where such interest areas and time provisions have been made, as in the World of Inquiry School in Rochester, New York, the results are impressive indeed.[1]

In summary then, the Piagetian conception of intelligence provides no support either for those who advocate formal preschool instruction or for those who argue for new methods and materials to stimulate intrinsic motivation. As we have seen, there is no evidence as yet for the lastingness of preschool instruction. In addition, intrinsic motivation seems best stimulated by allowing the child to engage in the activity of his choice for unbroken periods of time. As Jensen has so rightly pointed out, if we really want to maximize the effects of instruction, it does not pay to blink at the facts whether they have to do with racial or socioeconomic differences in intelligence, the effects of preschool instruction, or the nature of intrinsic motivation.

[1] The results of our preliminary evaluation of this school suggest that World of Inquiry pupils are significantly higher in their need for achievement and more positive in their self evaluations than are their matched controls (children taken from the waiting list) who are attending other schools.

References

Berlyne, D. E. Curiosity and education. In J. D. Krumboltz (Ed.) *Learning and the educational process.* Chicago: Rand McNally, 1965, 67-89.

Bertalaffny, Ludwig von. *Modern theories of development.* New York: Harper & Bros. (Torchbook Ed.) 1962.

Bloom, B. S. *Stability and change in human characteristics.* New York: John Wiley & Sons, Inc., 1964.
Bruner, J. *The process of education.* Cambridge, Mass.: Harvard University Press, 1962.
Burt, C. *Mental and scholastic tests.* London: Staples Press, 1962 (4th edition).
Burt, C., & Howard, M. The relative influence of heredity and environment on assessments of intelligence. *British Journal of Statistics Psychology,* 1957, **10**, 33-63.
Elkind, D. Quantity conceptions in junior and senior high school students. *Child Development,* 1961, **32**, 551-560.
Elkind, D. Piaget and Montessori. *Harvard Educational Review,* 1967 (Fall) 535-545.
Elkind, D., Barocas, R., & Rosenthal, B. Combinatorial thinking in children from graded and ungraded classrooms. *Perceptual and Motor Skills,* 1968, **27**, 1015-1018.
Fowler, W. The effect of early stimulation in the emergence of cognitive processes. In R. D. Hess & Roberta M. Meyers (Eds.) *Early Education.* Chicago: Aldine Press, 1968, 9 36.
Glick, J., & Wapner, S. Development of transitivity: Some findings and problems of analysis. *Child Development,* 1968, **39**, 621-638.
Goodenough, Florence. New evidence on environmental influence on intelligence. *Yearbook of the National Society for the Study of Education,* 1940, **39**, 307-365.
Goodnow, Jacqueline J. Problems in research on culture and thought. In D. Elkind and J. Piaget (Eds.) *Studies in cognitive development.* New York: Oxford University Press, 1969, 439-464.
Goodnow, Jacqueline J., & Bethon, G. Piaget's tasks: The effects of schooling and intelligence. *Child Development,* 1966, **37**, 573-582.
Hunt, J. McV. *Intelligence and experience.* New York: The Ronald Press, 1961.
Hunt, J. McV. Intrinsic motivation and its role in psychological development. In D. Levine (Ed.) *Nebraska symposium on motivation.* Lincoln: University of Nebraska Press, 1965, 189-282.
Jones, H. E. The environment and mental development. In L. Carmichael (Ed.) *Manual of child psychology.* New York: John Wiley & Sons, Inc., 1954, 631-696.
Kittlewell, H. B. D. Selection experiments on industrial melanism in the lepidoptera. *Heredity,* 1955, **9**, 323-342.
Montessori, Maria. *The Montessori Method.* New York: Schocken, 1964 (first published in English, 1912).
Moore, O. K. Orthographic symbols and the preschool child: A new approach. In E. P. Torrence (Ed.) *Creativity: 1960 proceedings of the 3rd conference on gifted children.* Minneapolis: University of Minnesota, Center for Continuation Studies, 1961.
Piaget, J. *Play, dreams and imitation in childhood.* New York: Norton, 1951.
Piaget, J. *Les relations entre l'affectivité et l'intelligence dans la developpement mental de l'enfant.* Paris: C.D.U., 1954 (mimeographed and bound lectures given at the Sorbonne).
Piaget, J. Genesis and structure in the psychology of intelligence. In D. Elkind (Ed.) *Six Psychological Studies by Jean Piaget.* New York: Random House, 1967a, 143-158.
Piaget, J. *On the nature and nurture of intelligence.* Address delivered at New York University, March, 1967b.

Piaget, J. Intelligence et adaptation biologique. In F. Bresson *et al* (Eds.) *Les Processus d'adaptation*, Paris: Presses Universitaires de France 1967c, 65-82.

Piaget, J. The mental development of the child. In D. Elkind (Ed.) *Six Psychological Studies by Jean Piaget*. New York: Random House, 1967d, pp. 3-73.

Rapaport, D. The theory of ego autonomy. *Bulletin of the Menninger Clinic*, 1958, **22**, 13-35.

Shaffer, L. F. Children's interpretations of cartoons. *Contributions to Education*, No. 429. New York: Teacher's College, Columbia University, 1930.

Skeels, Harold M. Adult status of children with contrasting early life experiences. *Monographs of the Society for Research in Child Development*, 1966, **31**, 3, No. 105.

Spearman, C. *The nature of "intelligence" and the principles of cognition*. London: Macmillan, 1923.

Strodtbeck, F. L. The hidden curriculum of the middle class home. In H. Passow, Miriam Goldberg and E. J. Tannenbaum (Eds.) *Education of the disadvantaged*. New York: Holt, Rinehart & Winston, 1967, 244-259.

Waddington, C. H. *The nature of life*. New York: Atheneum, 1962a.

Waddington, C. H. *How animals develop*. New York: Harper & Bros. (Torchbook Ed.), 1962b.

Wechsler, D. *The measurement of adult intelligence*. Baltimore: Williams & Wilkens, 1944.

Wechsler, D. *Wechsler intelligence scale for children*. New York: Psychological Corporation, 1949.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, **66**, 297-333.

Heredity, Environment, and Educational Policy

LEE J. CRONBACH, *Stanford University*

Professor Cronbach accepts some but by no means all of Professor Jensen's empirical conclusions. In the following review he indicates some research that bears on their points of disagreement. Cronbach suggests that such distinctions as Jensen's dichotomy between "Level I" and "Level II" abilities over-simplify the many dimensions of individual differences, and he disagrees with the educational policy he feels is implied by Jensen's recommendations for education. Beyond this, Professor Cronbach poses a more basic question—"Intelligence for what?"—a question of the compatibility of current social aims of schooling with long range changes in our social and technological structure.

Professor Jensen is among the most capable of today's educational psychologists. His research is energetic and imaginative. In the present paper, an impressive example of his thoroughness, I am sure every reader has had my experience of encountering valuable information in areas where he thought himself *au courant.* Unfortunately, Dr. Jensen has girded himself for a holy war against "environmentalists," and his zeal leads him into over-statements and misstatements. Rather than list the points where Dr. Jensen and I agree, and those where we diverge, let me begin with an integrated statement of my view on the major themes.

I do not doubt that performance—intellectual, physical, or social—is developed from a genotypic, inherited base. The organism, as it evolves prenatally and post-

* This comment was prepared under support from the U. S. Office of Education, but the views are those of the writer only.

Harvard Educational Review Vol. 39 No. 2 Spring 1969

natally, incorporates energy and information. What the person does with an experience, and what it does to him, depends on physical structures that were laid down during the previous years, or days, of his existence. Human development is a cumulative, active process of utilizing environmental inputs, not an unfolding of genetically given structures (Caspari, 1968).

The genetic populations we call races no doubt have different distributions of whatever genes influence psychological processes. We are in no position to guess, however, which pools are "inferior." Such a comparison is not meaningful, except in terms of the probability that the member of the group will be able to cope in some specific way with some specific challenge, after he has developed for a specified period in some specified environment.

Darwin's catchy phrase, "survival of the fittest," has misled hereditarians for a century. A genetic factor that has survival value in one environment is detrimental in another. Whatever the individual's genotype—barring gross defects—there are environments in which he will develop so as to function well and others in which he will develop poorly. For another genotype these effects may be reversed. There are many possible educational and developmental environments. At any age, the person is the phenotypic product of his genotype and his experiences to date; this phenotype may make him unready to profit adequately from the treatments now established for persons at that age. He might, however, be well equipped for some other series of educational procedures we could devise.

The phrase "improve the environment," born of the enthusiasm of the Social Darwinists, has misled environmentalists for two generations. Environments cannot be arrayed from good to bad, rich to poor. The highly stimulating environment that most of us think of as "rich" promotes optimal growth for some persons and may not be suitable for others. Environments can be varied along many dimensions, and the optimum with respect to each dimension depends on the person's phenotype at a given time. We think of the infant as deprived when he has nothing to gaze upon but a blank ceiling, but nothing is gained by making the environment so richly patterned that he cannot direct his attention. The pattern that holds attention varies with his age and his past experience (Fantz, 1961). An information-laden environment is rich, in some sense; but the right amount of redundancy and of detail depends on the learner's maturity. Conditions that make more information available may create an overload and so impair learning (Wicklegren and Cohen, 1962). How much stimulation is optimal, how much assistance, how much external monitoring and reinforcement, how much pressure for excellence, how much of the conceptual, and how much of the concrete—

these depend on the state of the individual and indirectly on his genotype. The optimum might be genetically determined; one can imagine, for example, metabolic differences that would make some children more impulsive than others and hence in need of a more calming environment. But the present state embraces biological structures, habits, attitudes, and meaning systems that are the residue of a long series of transactions. Some of them are transactions of genetically-determined structures with the environment, but more are transactions of the phenotype at a given moment with the environment of that moment.

There *has* been too much blithe optimism about our ability to improve the intellectual functioning of the slum child and the retarded child. Programs of compensatory education seem to have had no reliable and lasting effect. It may have been a sound political decision to launch massive compensatory programs, if only as a token of public concern. But far more was promised than we know how to deliver, and the hectic effort has drawn energies away from the needed basic, objective research. We need to clarify the aims of the programs and the hypotheses on which the experimental programs operate; this will move us beyond argument as to whether disadvantaged children can be helped, toward tested, workable plans (Hess & Bear, 1968).

There are two ways in which altered environment can be helpful, and these have been confused. One is to provide an optimal *maintenance* environment. That is how we promote the best growth of a plant; we select its planting spot, fertilizer, and so on to fit its requirements and maintain them throughout its life. This is likewise our way of overcoming the deficiency of the PKU child; we modify the environment by means of a special maintenance diet. The other is to provide a special environment for a brief *intervention* period, in the hope that the person will be brought to normal readiness for the conventional environment so that the special treatment can be discontinued. Remedial intervention can supply needed skills, alter habits, and overcome many physical defects. Which of the two lines of attack is suited for the problem of inaptitude for the existing educational environment is in part an empirical question.

But much of the decision rests on little-recognized policy issues. The intervention treatment intends that pupils placed in it shall ultimately complete the regular program of socialization and be indistinguishable as a group from those who did not need special treatment. I am sympathetic with the objection of Gordon and Wilkerson, quoted by Jensen, that it is wrong-headed to try to make the slum child fit the middle-class stereotype, as child or adult. But education must have a clear idea of its intended product. If we are to bring these chil-

Heredity, Environment and Educational Policy
LEE J. CRONBACH

dren to a self respecting adulthood, we must define for them a prospective role that has at least as great a value, to the individual and to society, as the middle-class model of industry, articulateness, social and cultural concern, and self-regulation No one protesting against middle-classness has gone on to describe a possible, viable society in which large subsegments of society have radically different orientations and functions.

Today's discontent is a clamorous crisis that distracts us from a quieter, yet more ominous crisis—the bankruptcy of long-range social planning. Lacking visions of what society might become, we are training people for a *status quo* that is already vanishing. The schools are committed to training people for production, responsibility, creation, and leadership. The intervention programs seek to offer that way of life to all. But the fact is that automation, centralization, complexity, and abundance already have created a society where most people work less and less, while the manager and the professional work 50 to 70 hours per week. Huxley's beehive *World*, where a few highly educated persons, conditioned to self-denial, carry the productive burden, is already the American way of life. It is against that world—where the uptight Alphas are the slaves—that our brightest youth are protesting. The time has come for far less concern with the total man-years of education produced by our system, and for intensive and sober concern with the capital question, "Intelligence for what?"

It is hard to see how evidence on heritability provides a base for social policy. It is surely humane policy—without regard to questions of heritability—to facilitate birth control. It is inconceivable that we will scale welfare payments to penalize the child in a large family. I hope it is inconceivable that data on heredity—whether of the individual or the group—will persuade us that some children should be taught concepts, some taught rote verbal associations, and some taught how to change tires. Jensen seems to argue that the disadvantaged should be taught by rote methods. But the cut-and-dried answers that can be learned by rote are not the answers that one needs if he is to cope with a changing world and to live an appreciative and expressive life. The proper and necessary strategy is to find alternative means of bringing all children as far as we can toward self-fulfillment. Under our present conception of the good life we cannot set goals of entirely different character for different pupils. It is regrettable that Jensen says little about the policies he would expect us to follow if we accepted his empirical conclusions.

On the scientific side, it is vital to break away from such stereotyped terms as "intelligence" and "learning ability." There is a spectrum of performances,

ranging from crystallized, overlearned routines to fluid information-processing, often referred to as *g* (Cronbach, 1969). Fluid ability is measured in tests like the maze, the matrix or figure analogies, block design, and embedded figures; Jensen's "Level II" abilities involve it. Crystallized abilities are diverse and specific: spelling of *-gn* words, handling of subjunctive clauses, etc. In schools as they now are, success is best predicted by taking inventory of the relevant crystallized abilities with which the pupil starts the year. The verbal "intelligence" test succeeds as a predictor primarily because it reflects concrete achievements. A child with average fluid ability and low crystallized ability is likely to do poorly; we have never succeeded in devising a mass educational program in which such a child is likely to achieve average success. Analytic ability should be a resource on which education builds, and as of now it is not.

Because learning abilities are plural, they are not adequately conceptualized by Jensen's Level I-Level II system. Many processes contribute to effective learning; some are under conscious control or trainable, and some not. Which processes are required depends on what is being learned and what kind of instruction is employed. At times, striking differences in "ability" can be overcome very simply. Lower-class children are inferior in paired-associate learning, according to many studies—no doubt the task would have a fairly substantial heritability (H) index.[1] But simply coaching the lower-class children to make up "meaningful" associations for the word pairs brought them up to the middle-class rate of learning. This finding, coming from Jensen and Rohwer (1965), seems to contradict what Jensen says in this paper. It is at least possible that on the Glasman "Level II" task (recall of objects that can be conceptually clustered) the lower-class children could overtake the middles if made aware of the usefulness of analysis. Indeed—stop the presses!—a brand-new study seems to demonstrate cleanly that very simple instruction does overcome initial weakness on the Glasman task (Moely *et al.,* 1969). Capability is not at issue when a child does not call upon an ability he possesses.

As to heritability, there is less here than meets the eye. The term, though standard in genetics, is mischievous in public discussion, for it suggests to the unwary that it describes the limit to which environmental change *can be* influential. Not so. The H index describes a certain population, having a certain gene pool and having developed in a certain range of environments. (There are some

[1] Heritability is "a population statistic, describing the relative magnitude of the genetic component (or set of genetic components) in the population variance of the characteristic in question." (Jensen, p. 42). Cf. Jensen pp. 42-43 for a formal definition of the term—ed.

treacherous assumptions, well explained by Huntley, 1966; the most critical of these is that environments are distributed at random over the various genotypes. But even if one made alternative assumptions, the H value would surely remain above 0.50, and it is hard to see how moderate changes in the index would alter one's social policy.) A phenotype that is 100 percent heritable is not affected by the variations among existing environments. But introduce a "mutant" environment, and H will change; this is exactly what happened in the case of PKU—a direct genotype-phenotype link was broken. Likewise, note the report of Osborne and Gregor (1966) that a certain type of spatial test has an H value of 0.89, alongside the finding (Brinkmann, 1966) that as soon as someone made an effort to train for this kind of ability, scores were increased by large amounts. The influence of environment on a trait with high H is also dramatically apparent to the American Fulbrighter of average height who finds large numbers of today's Japanese youth towering over him. Pool the heights of 1940 Japanese and 1970 Japanese in a single calculation, and H will be quite a bit lower than it is in either group alone. In most cultures mental-test scores show similar generation-to-generation gains attributable to environment.

Attention should be directed to Jensen's remark that environment does not affect stature above the level that includes "minimal daily requirements of minerals, vitamins, and proteins" (p. 60). But that is the point. Nutritional science now tells us to include certain chemical substances in the diet; these give heredity something to work with. You do not increase stature two inches in a generation just by providing more and more rice. You do not increase mental ability just by providing more stimulation. Analytic research will in due time specify the needed ingredients in an educational diet.

The heritability index of 0.80 is impressive, but it is less discouraging than Jensen implies; environments of the sort we now have can improve ability, if we can choose the environment to fit the individual instead of relying on fortuitous correspondences. A brief technical sortie will put a new light on the index of 0.80. Think of an "expected IQ"—the hypothetical average IQ of a thousand persons having identical genes, who have been assigned at random to environments. Assume that all IQs are "true" scores, perfect measures. Starting with Jensen's H value, the correlation of individual IQ with expected IQ, over the population, is $(0.80)^{1/2}$. The standard error of estimate for individual IQ is approximately $[200(1-0.80)]^{1/2}$ or 6.3. Hence persons having the same genes are distributed over an IQ range of more than 25 points. With run-of-the-statistics cases and within the range of present environments, the individual who draws an

environment fitted to his genotype develops an IQ some 6 points better than the expected IQ for that genotype, and 12 or more points better than does one who is unlucky in the draw. If an effect of this size could be brought under control and applied population-wide, it would surely be economically and culturally beneficial.

It is necessary to deal summarily with a number of aspects of the paper. I have detected substantial distortions in Jensen's report of some research, and I must therefore warn the reader against accepting his summaries. Selective breeding studies are a case in point; Jensen says that "maze learning ability" can be bred (p. 30). But Anastasi, interpreting the same data, emphasizes that the superiority of the selected stock was *not* due to any superior "learning ability" (1958, p. 91). In fact, some of the studies were carried out precisely to demonstrate that breeding selects on particular temperamental traits that facilitate learning under one condition and impede it under others. The maze learning superiority of the Tryon strains was specific to one kind of maze under one kind of incentive. I particularly invite the reader's attention to John Paul Scott's eloquent attack on the idea of a general inherited learning ability (see Rosenblith and Allinsmith, 1966, pp. 54-57), since Jensen cites Scott as if Scott endorsed such an idea (p. 30).

Jensen is severe with studies that encourage the belief that the retarded and the disadvantaged can be helped. He is right to be critical of many of the studies that claim positive results. He could well have cited the Zigler-Butterfield (1968) demonstration that simple increases in motivation for the test account adequately for most reported before-and-after differences in preschool children. He could justly have been more severe in disposing of the Rosenthal-Jacobsen study—which purports to find evidence that giving the teacher mental-test data biases the teacher's handling of the pupil. He gives excellent advice (pp. 96 ff.) on the design of evaluative studies. So far, so good. But when he cites the Wheeler study of a gain in IQ following the opening of the Tennessee hills to the modern world, around 1930, he goes out of his way to say, "The decline in IQ from age 6 to age 16 was about the same in 1940 (from 103 to 80) as in 1930 (from 95 to 74)." More accurately, the 16-year-olds declined from 95 at age 6 in 1930 to 80 in 1940. These adolescents were dropping behind the norm group, most likely because their schooling was not up to that of the norm group. Jensen notes (p. 17) that Bloom summarizes age-to-age correlations of mental-test scores. It seems to me that, having introduced this source, Jensen was obligated to disclose that Bloom gives these data an interpretation opposite to Jensen's. Bloom sees the gains from year to year in test scores as random and unpredictable, hence due to external

Heredity, Environment and Educational Policy
LEE J. CRONBACH

events and not inheritance. (This is one of several alternative interpretations that fit the data.)

There is plentiful evidence that late blooming occurs, i.e., that some persons rise dramatically in their relative position even as late as adolescence. To label these important effects as a "regression" effect (p. 99) neither explains nor diminishes them. There is a trivial regression effect, arising from sheer error of measurement on the earlier test; but most reports on large IQ changes indicate that the relatively low initial status was confirmed by several tests. Whenever prediction is imperfect because something has really happened between pre-test and post-test, there is some tendency for regression toward the mean, but that is no more than a paraphrase of the obvious: likely events happen more often than rare events.

Jensen accuses writers on education of underplaying or denying the role of heredity. Some of this bias does exist, but Jensen is unfair. He does not quote the writers in psychology and education who do devote space to heredity. And he does not see that, in writings for educators, it is pointless to stress heredity. The educator's job is to work on the environment; teaching him about heredity can do no more than warn him not to expect easy victories. Heritability of individual differences is not his concern. Even if, after education, rankings in ability were to correlate *perfectly* with some measure on the pupil's ancestors, the educator ought to be providing the best possible instruction he can for every pupil he faces. To be sure, the educator who makes policy has to decide, in allocating resources, whether to put more resources on the laggards or on the leaders, but this decision has to be based on a judgment about utilities. The same considerations enter this judgment, whether we assume zero heritability or perfect heritability.

Let me be telegraphic in disposing of some further reservations. Jensen states that "while fluid intelligence attains its maximum level in the late teens and may even begin to decline gradually shortly thereafter, crystallized intelligence continues to increase gradually with the individual's learning and experience all the way up to old age" (p. 13). I do not believe there is adequate evidence to offer a conclusion as to the trend of fluid ability with age. On another point, Jensen protests that we should not "reify *g* [general intelligence] as an entity" (p. 9), but it seems to me that he does so, especially as he begins to insist that it is a "biological reality" (p. 19). Fluid ability is demonstrated through a complex set of acts: attending, analyzing, encoding, transforming, etc. The process is not unitary even though the processes tend to be acquired through the same activities

and, so, to be correlated. Later Jensen concurs in Zigler's criticism of "unbridled environmentalists" in whose writing "the concept of capacity is treated as a dirty word" (p. 29). But "capacity" *is* a dirty word, incapable of being given meaning and overwhelmingly capable of confusing discussions. It and all words like it refer to nothing but an expectancy under present circumstances. Intellectual capacity is continually being expanded by technological devices. Perhaps Zigler and Jensen will protest that the computer has not really increased man's mathematical capacity (though he now can solve in a day problems that once took a lifetime). Do they not admit that the long-ago invention of a spoken language increased "capacity"? And if so, where can they draw a line?

Jensen does not present clearly the important concepts of covariance and interaction (pp. 38 ff.). Covariance exists whenever persons of a certain genotype experience anything other than a random selection of the environments; nothing about matching "good" environment to "good" heredity is implied. Interaction exists when a difference in treatments produces one difference in outcomes in persons of one genotype, and some other difference in outcomes with a second genotype. Jensen offers as an example of interaction the possibility that genetically different individuals will gain different amounts of weight when given the same number of calories. "Their constitutions cause them to metabolize the same intake differently" (p. 40). This is a poor example. One might paraphrase to say that Jensen thinks children who inherit good *g* "metabolize exactly the same environmental intake quite differently"—but his calculations take *that* as a main effect of heredity.

Finally, Jensen denies that there is severe deprivation in the home of the slum child (p. 61). I am no authority in these matters, but I have heard descriptions—e.g., of small girls locked into an apartment to keep them from the dangers of the street—that seem to qualify as severe deprivation. Bronfenbrenner (1967) asserts that the presence of severe deprivation in Negro homes is "an unwelcome but nonetheless inexorable reality." In addition to "the indifference and hostility of the white community," he believes that the child-rearing practices of American Negroes are "stubborn obstacles to achieving quality and equality in education" (p. 910). Jensen himself defines brilliantly a large part of what a child must learn before he is ready to participate effectively in presentday schooling (p. 7); the Negro child is often not given that training (Hess & Shipman, 1965).

It will be apparent that Dr. Jensen and I agree on many fundamentals. With regard to policy, we both believe that every intervention program has to stand on its demonstrated merits. I would not ask that it "raise the IQ," but I would

ask that it raise readiness for schooling or promote intrinsically valuable achievement. We both urge that new kinds of instruction be devised to fit diverse patterns of ability. One goal of instruction, in my opinion, should be to develop fluid ability and conceptual learning ability. The undoubted significance of heredity must not deter researchers from trying to design procedures that will do this. Impossible things are happening every day.

References

Anastasi, A. *Differential psychology* (3rd edition). New York: Macmillan, 1958.

Brinkman, E. H. Programmed instruction as a means of improving spatial visualization. *Journal of Applied Psychology*, 1966, **50**, 179-184.

Bronfenbrenner, U. The psychological costs of quality and equality in education. *Child Development*, 1967, **38**, 909-925.

Caspari, Ernst. Genetic endowment and environment in the determination of human behavior: Biological viewpoint. *American Educational Research Journal*, 1968, **5**, 43-56.

Cronbach, Lee J. *Essentials of psychological testing* (3rd edition). New York: Harper & Row, 1969. In press.

Fantz, R. L. The origin of form perception. *Scientific American*, 204:5, May, 1961, 66-72.

Hess, Robert D. & Bear, Roberta M. (Eds.). *Early education*. Chicago: Aldine, 1968.

Hess, Robert D. & Shipman, Virginia C. Early experience and the socialization of cognitive modes in children. *Child Development*, 1965, **36**, 869-886.

Huntley, R. M. C. Heritability of intelligence. In J. E. Meade & A. S. Parker (Eds.). *Genetic and environmental factors in human ability*. Edinburgh: Oliver & Boyd, 1966, pp. 201-218.

Jensen, Arthur M. & Rohwer, W. D., Jr. Syntactical mediation of serial and paired-associate learning as a function of age. *Child Development*, 1965, **36**, 601-608.

Moely, Barbara E., Olson, Frances A., Halwes, Terry G., & Flavell, J. H. Production deficiency in young children's clustered recall. *Developmental Psychology*, **1**, 1969, 35-39.

Osborne, S. T., & Gregor, A. J. The heritability of visualization, perceptual speed, and spatial orientation. *Perceptual and Motor Skills*, 1966, **23**, 379-390.

Rosenblith, Judy F., & Allinsmith, Wesley. *The causes of behavior* (2nd edition). Boston: Allyn & Bacon, 1966.

Wicklegren, W. & Cohen, D. H. An artificial language and memory approach to concept attainment. *Psychological Reports*, 1962, **11**, 815-827.

Zigler, E. & Butterfield, Earl C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. *Child Development*, 1968, **39**, 1-14.

A Letter from the South

WILLIAM F. BRAZZIEL, *Virginia State College*

This letter came, in its earliest form, even before subscribers had received their copies of the Winter issue. Mr. Brazziel had been following the Jensen controversy ever since Dr. Jensen spoke to the American Educational Research Association in 1968. When news of the Harvard Educational Review *article reached Mr. Brazziel through local publicity (newspapers and the coverage in* U. S. News and World Report) *the correspondence printed below began. In subsequent issues of the* Review *we expect to print further letters and comments from our readers.*

Sirs:

Thirteen years ago plaintiffs brought suit in Federal District Court to integrate the Louisiana public schools. The main argument of the defense attorneys and the superintendent of public instruction was that "white teachers could not understand the Nigra mind" and, therefore, would not be able to teach them effectively in integrated classrooms. The defense quoted heavily from the theories of white intellectual supremacy as expounded by Henry Garrett and Audrey Shuey

Last week, a scant five days after Arthur Jensen made headlines in Virginia papers regarding inferiority of black people as measured by IQ tests, defense attorneys and their expert witnesses fought a suit in Federal District Court to integrate Greensville and Caroline County schools. Their main argument was that "white teachers could not understand the Nigra mind" and that the Nigra children should be admitted to the white schools on the basis of standardized tests. Those who failed to make a certain score would be assigned to all black remedial schools where "teachers who understood them could work with them." The defense in this case quoted heavily from the theories of white intellectual supremacy as expounded by Arthur Jensen.

Harvard Educational Review Vol. 39 No. 2 Spring 1969

Letter from the South
WILLIAM F. BRAZZIEL

It will help not one bit for Jensen or the HER editorial board to protest that they did not intend for Jensen's article to be used in this way. For in addition to superiority in performing conceptual cluster tricks on test sheets, the hard line segregationist is also vastly superior in his ability to bury qualifying phrases and demurrers and in his ability to distort and slant facts and batter his undereducated clientele into a complete state of hysteria where race is concerned.

Jensen and the HER editorial board will modestly admit that they have superior intellects and I am sure they realized the consequences of their actions. Questions now arise as to why they decided to raise this issue, in this way, and at this time.

Fortunately, doubts about the ability of black and yellow people to master war, finance, science and technology are waning rapidly in both white and black minds. The imprecision of standardized testing is now clear to most literate people and the criminal use to which they are put in schools is also becoming clearer. Black history has made people aware that white people did *not* give America such things as the stoplight, the shoe last, heart operations and sugar refining but that black people did this. That John Smith did not develop corn and tobacco but learned to grow these crops from the Indians. And the beat goes on. People are now witnessing with their very eyes the fact of black youth finally given a half of a chance at education and jobs and being able to make exotic formulas for bombs and napalm as well as anyone else. As a result of all of this, I think the present set-to might be the last go-round for white supremacy psychological theory.

I would hope the Jensenites could alter their stance and approach and try to bring some good out of this situation after all. They might work their way out of ethnic learning styles by broadening their research to include all ethnic groups. We have some rather learned men in our area who believe that English-Americans are atop the pyramid of abstract learning abilities with Welsh, German, French, Belgian, Norwegian, Swiss, Finnish, Danish and Swedish occupying the next nine rungs in the order listed. After the top ten have been given their just due, these gentlemen give a smattering of attention to the rest of Europe and proceed to ignore the rest of the world. The Jensenites might try to clear this up in some way. They might even look into intra-group differences within the top ten. I would suspect that many would be found and that it would be healthy to make this known at professional meetings, in the journals and in the news media.

We also have a religious wing in this group who suspect that English-American children who are brought up in Southern Baptist churches perceive things differently and might really deserve the top spot upon the pyramid. Southern-

English-American-Episcopalians regard these assertions with a great deal of amusement. But who really knows? We all will if the ethnic learning line of research is extended logically to include every possible ethnic, regional and religious stock.

Also in the status research vein, we need research on the effects of racism and caste status on learning. The Jensenites can provide this by following Robert Coles and others around in Mississippi and South Carolina to study the parasitic worm and starvation situation among black children. Autopsies of a few who died might yield valuable evidence on the brain damages wrought by malnourishment. The team could change themselves into black people ala John Griffin and run the hostility gauntlet as they tried to find some information in the local library. Or the hilarity gauntlet as they made application for a professional or skilled job. They could fly as black men to Boston or Oakland and make the same applications to the craft union nearest the airport. Or they could try to get a tenured appointment in the Harvard Graduate School of Education, or a spot on the HER editorial board, or simply a rank higher than assistant professor among the 7,000 member Harvard faculty.

The Jensenites could give the same black injections to their children, enroll them in a different school and record what happens to them. Children learn efficiently if listening, reading, discussion, peer-group interaction, library resources and teacher-pupil interaction are all used efficiently. The investigators might be very interested in the change in quality in the last four areas for their now black off-spring and to see who is to blame and how the situation can be improved. To add a spicy dimension, low IQ scores could be substituted in the transfer folders.

Creation of multi-ethnic and multi-racial tests would also be a method of bringing some good out of the situation. If the only way to make *exactly* the same score on test items is to be of the same race, economic class, ethnic stock, and religious persuasion as the committee that developed the instrument, then we either must make intensive efforts to inter-marry, re-distribute income and institute religious purges and programs in this country or we must try to integrate more multi-racial and multi-ethnic material into the instruments. Said in the words of Dr. Nathan Wright, the Newark black power theorist, we must try to "dehonkify" the instruments.

Or we might decide that making *exactly* the same score is not important for all races and religions and come up with an Ethnic Success Quotient for tests based on validation studies of all of the hypenated groups we are going to study. Under such a system a Richmond born, Episcopalian, of English stock, from a family with an income of $12,000 would be declared below average if his Binet score

was below 120. A score of 100 would relegate him to success quotient oblivion as a low normal. The Beaufort County, S. C. black children with worms might have a success quotient of 90 based on performance of adults from this sort of situation who somehow scrambled up the ladder. A black 100 score in this county would indicate a ESQ of potential genius.

Finally, in this vein, the Jensenites might make their most important contribution if they could somehow join with Earl Schaefer of the National Institutes of Health and others at the Universities of Florida, Western Michigan, etc. who are fastening on early infant stimulation and teaching as the key to agility on standardized tests. (The problem, of course, may be in getting the Schaefcrites to join with the Jensenites given the Klan types who have embraced the latter as their own). Schaefer has already published some fine results of efforts with black children. The logic here is simple and very much in the vein of Cronbach's rebuttal to the Jensen paper, i.e., if you want black kids to think like white kids, imprint this type of thinking habit early (5 days to 2 years of age) with simple thinking, concept cluster tasks. White teachers can enable black parents to learn how. White disadvantaged children are being imprinted in the same manner in some studies. Ethnic and religious backgrounds have not been treated as yet. There might be a problem or two here regarding people who might want to imprint their children with their own brand of thinking or who have deep affection and preference for certain racial, ethnic or religious ways of thinking. Other parents might not want the new imprints to attend their schools on an integrated basis or live in their neighborhoods and play in their recreation centers. Something in the imprinting would thus be lost in this sort of forced isolation. But I am certain these reservations can be swept aside in the name of psychological research and the cognitive homogenizing process can progress.

Now for a closer look at some of Jensen's theories about black IQ. To begin, I received a form letter from Jensen in response to a request for clarification of his *real* stand on the implications of racial genetic inferiority that seemed to shine through the somewhat hazy statements of conclusion of his paper at the AERA, implications which the press quickly translated into flat statements of white intellectual supremacy. His article was based on this paper and gave the same impression to the press. (See Joseph Alsop, *Washington Post*, March 11; *Virginian-Pilot*, March 12: "Yet there is no use being mealy-mouthed about it. Dr. Jensen is really saying that in *addition* to the handicaps wickedly imposed by prejudice and discrimination, the average black American begins the race of life with a detectable genetic handicap").

Jensen's letter was addressed to the *Berkeley Daily Gazette* which he feels misinterpreted his position. The following are excerpts from the letter:

> Obvious differences in inborn mental ability 'between races'—these are a reporter's words. They certainly are not mine. The quotation marks, attributing this phrase to me, are therefore wrong. Furthermore, the statement is quite indefensible. The complex *causes* of objectively measurable differences in mental abilities among individuals or between different socioeconomic and racial groups are not at all 'obvious'.
>
> Although my study of the existing evidence has led me to the position that intelligence differences among individuals, social classes, and racial groups are conditioned by both genetic and environmental factors, the estimation of the relative contributions of these influences is a problem of great technical and practical difficulty for researchers in behavioral genetics, and the research so far has been inadequate as a basis for definitive conclusions about racial differences in intelligence.

Jensen's treatment of the racial aspects of IQ in his article comes to the same point of inconclusiveness. It is very, very unfortunate that he, or the editors, failed to include a clear statement to this effect. Truth squad operations such as this letter and the rebuttals by psychologists in the HER Spring issue will never get read.

Jensen's second error in my estimation was to lean heavily on the Coleman Report for data on black inferiority. This report has been heavily criticized for inaccuracy. The most notable criticism is contained in the Winter, 1968 issue of the Journal of Human Resources in an article by Bowles and Levin. Sampling procedures, lack of cooperation by big school systems, failures to match black-white sample by curriculums, over-reliance on administrators' contentions that black-white facilities were indeed separate but equal (black parents in Eutaw, Alabama must have thought the research team had been smoking pot when they read the conclusions of the report) and crudeness of statistical measures were all analyzed as weaknesses which, when added to the fact that the study was made in pre-ESEA days, relegated it to the status of a 737 page, million dollar pilot study. On page 292 of the report, the authors state similar disclaimers, especially regarding the precautions necessary in interpreting their statistics.

In regarding as law this report's conclusions that the average black kid can get no further than a 9th grade operating level after 12 years of public school, Jensen ignores completely (or is unaware of) the record being compiled by the JOBS program of the National Alliance for Businessmen. These gentlemen take black drop-outs, place them on the job half-time and in reading and math classes half-time; they produce a two-year gain on tests every six weeks.

Letter from the South
WILLIAM F. BRAZZIEL

Jensen's major error, I believe, was his inconsistency in following a definite line of reasoning regarding the separation of gene linkage and pre-postnatal ravages of protein malnutrition. The latter is the most intensively researched thesis these days with NIH teams leading the way. Jensen did not even mention this line of research which (together with research in infant stimulation) I believe has answers for 42% mental retardation found in low-low (Jensen's level V) income black children and a lot of the other differences. In a half-starved brain like these kids have, how are we to really know if high or low IQ genes were linked? Jensen did not tell us how.

Jensen calls compensatory education a failure. So did reporters of the *Washington Post* who in turn received and printed a report by the ESEA staff of the Virginia Department of Education calling their allegations inaccurate and stating that they had hard data to back their claims. In response to a request for same, I received tables for statewide pre-post testing of 10,200 pupils in 15 school districts for 1967-68. The data show average month's increase in grade equivalency per month of 1.06 of instruction or an average overgain in achievement of more than a half a year per pupil as a result of compensatory education. Children scoring in the lowest decile had decreased from 41% to 28%. In the second quartile the number jumped from 8% to 16% and the drop-out rate had decreased by 63%. The officials noted that age-grade decrement had been scotched and that they believe that they had convincing evidence that their Title I program was a success. And this from one of the more conservative states in the Union and one with a record of slow starts in educational innovations. School people, it seems, are just now learning how to run compensatory programs. Or really try to. The first report to the President of the National Advisory Council on Disadvantaged Children noted this reluctance to really plan and implement on the part of many school systems. They quoted one superintendent who stated flatly that "it was useless and a waste of money to teach those jigs anything." Let us all hope he has since initiated a good program and that he doesn't read Jensen's article.

In drawing conclusions from 200-300 comparative studies of black-white IQ, Jensen failed to consider that all of the pre-1948 studies and most of the post-'48 studies failed to give attention to the deprivation axioms made popular by the University of Chicago group (Davis, Eels, et al) and until recently almost no psychometrists gave attention to the fact that white examiners in a black classroom are, in many, many cases, getting an invalid test performance. Their color, voice, manner, gestures turn many kids off, and they refuse to try. *This phenomenon is growing in intensity and must be dealt with.* How are you going to have a

77-942 O - 72 - 43

valid test session with kids who read in black papers and magazines that white researchers are sending their kids tc Harvard by over-studying the black communities with federal grants? Or with kids who received a leaflet from a community group blasting tests as an "unfair tool of colonialists who control the black community"?

I believe that Jensen is wrong and I hope he does not do too much damage. I believe the HER editorial board should publish the rebuttals in the same issue with future attacks on the Negro. Rumors abound that attacks on the Negro church are planned. This will scotch the sensationalism of the press caused by the lag in time between issues. Indeed, the rebuttals will never be read by reporters, much less printed.

Jensen failed to take into consideration the black infant mortality rate as a factor in black infant supremacy on the motoric area of the Bayley Scales. This rate is three times that of white infants. Black kids must literally undergo a survival of the fittest test to be born, once conceived, and to stay alive.

Jensen has a serious contradiction in his analysis of tests and studies of black IQ. After offering half dozen or so studies to document his thesis that black kids don't do as well on IQ tests as white kids, Jensen closes his paper by stating that IQ tests fail to measure the full potential of black kids.

Jensen failed to consider the 1969 report of the Research and Evaluation Branch of Project Head Start in writing off Head Start gains as transitory. According to this report of several studies of the maintenance of gains, the investigators concluded that the gains were maintained when the children were enrolled in first grades or kindergartens in middle-class schools. Edmund Gordon of Teachers College and John McDavid of Miami led the team which wrote this report.

Jensen, like other psychologists, is completely incapable of un-raveling what would have to be un-raveled in order to separate genetic from environmental influences where American black and white people are concerned, to wit:

1. If 90% of the black people in America have ancestors that include white people, how can we tell when black genes or white genes make for a wrong mark on a test score sheet?
2. If a large per cent of white people have black ancestors, who are they? Are their samples controlled for this factor? Which genes, black or white, make for right marks on a test score sheet?
3. How can we parse out the effects of brain damage, brain stunting (due to malnutrition) and lack of early stimulation? Which accounts for a wrong mark on the test score sheet?

4. How can we parse and measure the degree of access and *welcome* of black people to cultural learnings?
5. How can we parse and measure the interest in and acceptance of the white "way of life" by black mothers and children? One can't get good scores on a "way of life" test like IQ unless one lives and accepts this life fully.
6. How can we develop indices which show comparability of school strengths, weaknesses and emphases? The school assessment study by Tyler's group is just getting underway over loud cries from many school people.

Jensen failed to consider the learning styles of black parents and the origins of these learning styles when he made white-black comparisons on associative and problem-solving learning. If you go to many rural schools in the south today, you will find the associative type of learning proceeding as it has for many, many years —*for both races.* This is the learning heritage of most big city black parents. They pass this style on to the kids early and it shows up in test profiles. If conceptual learning is viewed as a gradual acculturation process and offered early in school careers, these kids can be made to think. Jensen's exhortations to teachers to rely completely on associative learning might preclude this ever becoming a reality, however. Before any more articles are published, I think Jensen should do more work in the area of black history, demography and culture and that he should try to get into the area of racism and isolation and the big role they play in differences. There really is merit in his actually taking the black injections and getting first-hand information. He would only have to be a black man for two months.

Jensen's "*g* factor", the main basis of his claims for white supremacy, cannot be accepted as the mysterious phenomenon he postulates. Even little children now know from their television science that if something really exists, scientists will isolate it and measure it—especially before making serious conclusions about it.

I believe Jensen made two good points. One is that IQ tests don't show the full learning potential of kids who are poor and black. I was happy to learn that he had invented a test which does a better job. We should all buy it. He should make millions. The other is that intensive instruction rather than "cultural enrichment" is necessary to make these kids learn if they are locked in neighborhood schools. Unlike Jensen, I believe that they can proceed from associative learning to abstract reasoning if the instruction gradually brings them to this point. And even with this, I believe black kids will continue to think and score test items differently until full equality is achieved. Black kids screen out much of the curriculum and perceive the rest differently. Consider perceptions of Tarzan and the British Em-

pire, for examples. Of course some black nationalists feel that it is a blessing that black people don't think like white people. As long as they can handle modern technology, make war, manipulate stocks, etc., I don't guess it really matters.

I believe the most potent strategy in the end will prove to be a combination of early stimulation and imprinting, and integrated schools with teachers who are free of racial and social class prejudices. IQ tests will also be eliminated from the schools. This is the strategy on which Neil Sullivan based his cross-bussing operations for the Berkeley schools. This may account for some of Jensen's concerns and reservations and perhaps, for his article. Pettigrew and others presented evidence in their work for the Civil Rights Commission that the earlier black children were placed in integrated schools, the closer they came to white norms on achievements tests. In turn, the white children came closer to perfection in their social learnings while losing no ground in test proficiency. The black children pick up the mysteries of Jensen's "*g*-factor" through association, I suppose, while the white children pick up the mysteries of "soul."

Reducing the Heredity-Environment Uncertainty

ARTHUR R. JENSEN
University of California, Berkeley

In his comments on the seven responses Dr. Jensen replies to criticisms, and suggests some appropriate research endeavors that could provide answers to the questions raised in his original article.

When the Editors of the *Harvard Educational Review* invited me to write a comprehensive summary of my research and thinking on the subject of educationally relevant individual differences, with reference especially to their genetic basis, I was delighted for the opportunity to present my views to the diverse and sophisticated audience that is reached by this journal.

One of my purposes in writing "How Much Can We Boost IQ and Scholastic Achievement?" was to provoke discussion among qualified persons of some important issues I believe have been relatively neglected in our common concern with improving the education of children called disadvantaged. Therefore it is a source of great satisfaction to me that the Editors have solicited and received extensive discussions of my article from several distinguished psychologists and an eminent geneticist—men whose own research in a variety of fields most germane to the contents of my article is widely known and highly respected.

Points of Agreement

It is of interest that many of the reports of my article in the public press have tried to make it look as though the several commentaries solicited by the Editors are strongly opposed to my paper and are in marked disagreement with its main points.[1]

[1] *U. S. News & World Report* (March 10, 1969), *Newsweek* (March 31, 1969), *Science News* (April 5, 1969), *Time* (April 11, 1969).

In fact, seldom in my experience of reading the psychological literature have I seen the discussants of a supposedly "controversial" article (in the Editors' words) so much in agreement with all the main points of the article they were asked especially to criticize. On my main points the discussants agree with me at least as much as they agree among themselves, which is considerably.

The Role of Heredity

On this central theme there is essential agreement. Crow, the population geneticist, states: "That the heritability [of intelligence] is large is a justifiable conclusion at this stage. . ." "I agree with Jensen in deploring an uncritical assumption that only environmental factors are important and that genetic differences are negligible." "We should also realize that to whatever extent society is successful in its goals of providing equality of opportunity, to that extent the heritability [of mental abilities] will increase." Bereiter, a leader in psychometrics and in early childhood education, makes the same points: "The heritability of intelligence is unquestionably high, but what is more to the point is that with further social progress the consequences of heredity can only be more important because of the elimination of such sources of environmental variance as differences in the quality of education, nutrition, and medical care." Cronbach, our most eminent educational psychologist, says there is no doubt that "performance—intellectual, physical, or social—is developed from a genotypic, inherited base." Elkind, the leading American exponent of Piagetian psychology, emphasizes Piaget's agreement with genetic and biological maturational factors in cognitive development. Piaget's indices of cognitive development, such as the ability to conserve quantity, area, and volume, have been factor analyzed along with traditional psychometric measures of intelligence and are found to be highly loaded on the *g* (general intelligence) factor (Vernon, 1965); and Tuddenham (1968) has found social class and racial differences on a psychometrized form of the Piagetian developmental tasks that are comparable to those found for nonverbal IQ tests. Other supporting evidence relevant to this conclusion has been reviewed by Kohlberg (1968) in a paper highly germane to my own formulations. An interesting indication of the role of genetic factors in these Piagetian indices of cognitive development has recently come to my attention in a study by De Lemos (1966), who found that a majority of the full-blooded Australian aborigines who were examined on a variety of Piagetian conservation tests still did not show conservation of quantity, weight, volume, number, and area, even by the time they had reached adolescence. (The majority of European children pass these tests by seven years of age.)

These tests were passed, however, by a significantly larger proportion of aboriginal children who had one European grandparent or great-grandparent. De Lemos does not account for these results in terms of possibly differential environments. De Lemos's data are shown in Table 1.

TABLE 1

Numbers of Full-Blood and Part-Blood Australian Aboriginal Children Passing Piagetian Conservation Tests and the Significance Level (p) of the Difference[a]

	Age 8 to 11 Years			*Age 12 to 15 Years*		
	Full	*Part*	*p*	*Full*	*Part*	*p*
Total N =	25	17		17	21	
Tests						
Quantity	2	6	<0.1	2	15	<0.01
Weight	9	11	<0.1	7	17	<0.01
Volume	0	5	<0.05	2	4	N.S.
Length	10	10	N.S.	3	13	<0.05
Number	0	4	<0.05	3	8	N.S.
Area	1	4	N.S.	2	8	N.S.

[a] Source: De Lemos (1966).

Genetic Component in Race Differences

Here, too, there is considerable agreement, although it is qualified in some instances in ways that I will examine in later sections. In my paper I proposed simply that the hypothesis of genetic racial differences in mental abilities is a reasonable one deserving of further scientific investigation. Crow states: "I agree that it is foolish to deny the possibility of significant genetic differences between races. Since races are characterized by different gene frequencies, there is no reason to think that genes for behavioral traits are different in this regard." Cronbach agrees that "the genetic populations we call races no doubt have different distributions of whatever genes influence psychological processes." He then goes on to say: "We are in no position to guess, however, which pools are 'inferior.' " On this statement two comments are in order: First, who has advocated that we merely *"guess"* about racial genetic differences? I am advocating that we seek objective answers regarding genetic differences through appropriate scientific research. Again, the point I made

in my article was that the present evidence on this topic is such that the hypothesis of genetic racial differences in intelligence is not an unreasonable one and should therefore be the subject of scientific investigation. Second, why does Cronbach put quotation marks around the word *inferior*? Lest the reader incorrectly infer that Cronbach is quoting me, let me note that I myself do not use this term and I object to it in this general context. I have said that there are racial and social-class differences in *patterns* of abilities and that there are probably genetic as well as environmental factors involved in these differences. The terms *inferior, superior, high, low, above, below*, etc. are meaningless in psychological discussions unless some particular dimension in the whole realm of abilities or traits is clearly specified and its relevance to a particular environmental adaptation is understood. Cronbach knows as well as I that it is nonsense to speak of different racial gene pools in general as *superior* or *inferior*.

Possible Dysgenic Trends in Our Population

In my paper I raised the question: "Is there a danger that current welfare policies, unaided by eugenic foresight, could lead to the genetic enslavement of a substantial segment of our population?" Differential birthrates in the population that are correlated with educationally and occupationally relevant traits of high heritability could produce long-term dysgenic trends which would make environmental amelioration of the plight of the disadvantaged increasingly difficult.[2] Hunt, psychology's most eloquent and influential spokesman for environmental amelioration of educational handicaps, states that ". . .the national welfare policies we established in the 1930's have probably operated in dysgenic fashion, and that it is highly important to establish welfare policies which will encourage initiative and probably, in consequence, help foster positive genotype selection." Hunt points out how some social and educational programs, such as involving parents in programs of early childhood education, can produce not only direct benefits to the children enrolled in the program but also more indirect benefits to the future welfare of the families involved, as when parents voluntarily enrolled in a Planned-Parenthood clinic. Says Hunt: "The enrolling in the Planned-Parenthood clinic suggests that this kind of enterprise in early childhood education instigates help to prevent some of the dysgenic processes with which Professor Jensen and I are both concerned. Hunt also agrees that it is "highly important to raise the intelligence, the educational

[2] For instance, unless existing trends markedly change, it can be predicted that within the next 20 years more than a million children with IQ's below 70 will grow up in fatherless homes in our urban slums. The amount of human frustration and suffering implied by this prediction, if it becomes reality, is incalculable.

Reducing the Heredity—Environment Uncertainty
ARTHUR R. JENSEN

attainments, and/or the general competence of those people who now comprise the bottom quarter of our population in measures of this cluster of characteristics."

If Hunt believes there have probably existed dysgenic trends in some segments of the population since the 1930's, he must logically conclude that he also believes there are heritable behavioral differences among some segments in the population, socially and educationally relevant behavioral differences that exist within every racial group, although he does not say this explicitly. There is, of course, nothing "inevitable" about these genetic differences in the sense of their being predestined or immutable or inherently associated with race *per se*. Whatever they are, if they indeed exist, they are undoubtedly a product of differing historical, social, and environmental selective pressures. The really important point now is to try and understand the genetic trends in the population resulting from current social forces, and if dysgenic trends indeed exist, to discover the kinds of social conditions and public policies that can be created in a humane, democratic society to counteract and reverse such trends for the good of all, especially of the generations not yet born.

Value of Compensatory Education Programs

I am essentially in agreement with Hunt's evaluation of the failures of compensatory early childhood education and the reasons for the ineffectiveness of preschool programs based on the free-play socialization model of the traditional nursery school. One must also agree with Hunt that we cannot now evaluate forms of compensatory education that have not yet been tried or even invented. The fact remains, however, that our most massive, large-scale attempts at what has been called compensatory education have apparently not produced the desired or promised results. I cited the comprehensive evaluation of the U. S. Commission on Civil Rights (1967), which arrived at this negative conclusion after a nationwide survey of the major Federally-funded compensatory programs. I favor continuing experimentation in improving the education of the disadvantaged, and I favor trying a wide diversity of reasonable approaches. In our present state of ignorance about how best to teach children who are spread over an enormously wide range of abilities and proclivities and diverse cultural backgrounds, we are hardly justified in launching nationwide compensatory programs of massive uniformity. The same expenditures invested in a real *variety* of smaller-scale programs that psychologists, educators, and parents have some reason to believe might succeed, and which can be properly evaluated, will more surely and quickly lead to knowledge of which policies and practices will or will not produce the most beneficial results. We *have*

learned from many of the programs evaluated by the U. S. Commission on Civil Rights what kinds of measures have produced no signs of success, though they have been put to the test for from three to eight years. It is a half-truth to say that these programs have not had a fair trial. Thirty years after the beginning of the progressive education movement, its extreme proponents, then on the defensive, were still saying it could not be evaluated because it had not been tried for a sufficient time. At least from the evidence now at hand, I must agree with Cronbach's statement that there has been "too much blithe optimism about our ability to improve the intellectual functioning of the slum child and the retarded child." And Elkind says "What is the evidence that preschool instruction has lasting effects upon mental growth and development? The answer is, in brief, that there is none." Bereiter, on the other hand, presents new evidence from his own excellent work with disadvantagd preschool children showing substantial gains in intellectual skills resulting from specific forms of intensive instruction. These are exciting findings and we will want to follow this work closely in the future. The crucial question, we all recognize, still concerns the permanence of the gains and the factors that affect their durability. The answer is still in the future.

Points of Disagreement

The points of disagreement seem to me less fundamental and much narrower in scope than the points of agreement. Some of the most critical-sounding statements quoted so repeatedly in the public press actually have little if any substance to back them up when read in context. At least two of the discussants seem to disagree with each other regarding my objectivity and accuracy. Crow states: "Jensen's article, together with many others that he has written recently on this subject . . . , constitutes a thorough review and synthesis of the various attempts to apply these methods [of biometrical genetics] to human intelligence and scholastic achievement. Jensen has become a leader in this field, and I, as a population geneticist, admire his understanding of the methods and his diligence and objectivity in bringing together evidence from diverse sources. He presents the evidence fairly, relying on empirical data in preference to introspection or traditional wisdom, and is very careful to distinguish between observation and speculation." Cronbach, on the other hand, makes a highly contrasting statement in the first paragraph of his paper: "Unfortunately, Dr. Jensen has girded himself for a holy war against 'environmentalists,' and his zeal leads him into over-statements and

misstatements." Since this has become the most widely quoted critical statement in the press about my article, I would like to examine it.

Let readers judge for themselves if there is anything warlike about my article. There is little doubt, however, that in recent years students of the behavioral and social sciences, educators, and the public in general have been strongly propagandized with the views espoused by extreme environmentalists, and that these views have become a basis for official policies.[3] If Cronbach interprets my confronting those he refers to as "environmentalists" with some of the scientifically-ascertained facts concerning the genetic aspects of mental abilities as being a "holy war," that is interesting in itself. What Cronbach calls a "holy war" I call simply looking for the facts.

But what about the more serious allegation that Cronbach goes on to make—that of "over-statements and misstatements" in my article? Cronbach does not follow up on this charge. He does not point to a single example of an "over-statement" or a "misstatement" in my paper. The closest Cronbach comes to indicating specifically what he might have had in mind in using these words is later on, where he says: "I have detected substantial distortions in Jensen's report of some research, and I must therefore warn the reader against accepting his summaries. Selective breeding studies are a case in point . . ." Let's take a close look at how Cronbach follows up on this attempted broadside.

Selective Breeding Studies

I stated that rats can be bred for maze-learning ability. I also pointed out that maze learning is a complex behavior, involving a host of sensory, motor, temperamental, neurological and biochemical components. Nevertheless, the molar behavior of speed of learning to run through a maze without entering blind alleys, I said, can be selectively bred. Cronbach seemingly challenges my statement by pointing out almost exactly what I had already stated in my own paper, namely, that maze-learning ability is a result of many factors. One can breed for any particular pattern of these factors, depending on the nature of the learning task and the criterion which serves as the basis for selection in the breeding of successive generations. Cronbach notes that the Tryon strains were bred to one kind of maze

[3] We find, for example, a statement from the U. S. Office of Education (1966): "It is a demonstrable fact that the talent pool in any one ethnic group is substantially the same as that in any other ethnic group." And from a Department of Labor (1965) report: "Intelligence potential is distributed among Negro infants in the same proportion and pattern as among Icelanders or Chinese, or any other group." There is simply no factual basis for these official pronouncements, which I believe are motivated more by political than by scientific considerations.

under one kind of incentive. Is the selective breeding for maze learning in one highly specific set of conditions any *less* genetic than breeding for maze learning ability that generalizes across many different mazes? In fact, in the study which I cited as an example, and from which my Figure 4 is taken, rats were bred for learning ability that generalized across 24 different mazes. I would call this a fairly general factor of maze learning ability. Fuller and Thompson (1960) in their well-known textbook, *Behavior Genetics,* say of this experiment:

> Thus a fairly broad range of rat intelligence was sampled. The procedure involved a lengthy period of habituation for all animals on simple pretest problems until a certain criterion was reached. In this way, the influence of motivational and emotional differences was minimized. The Hebb-Williams maze, generally speaking, is analogous to human intelligence tests which involve a large number of short items usually administered only to subjects who have had previous preparation. (pp. 212-213)

Since the 1953 paper by John Paul Scott that Cronbach refers to as an "eloquent attack on the idea of a general inherited learning ability" predates the Thompson experiment to which I referred, the only maze learning experiments it cites being those by Tryon, who bred rats for a specific maze ability, it can no longer be regarded as an adequate account of what we now know about selective breeding for maze-learning ability. Indeed, I have found no evidence in the literature of a general learning ability factor in animals that generalizes across a wide variety of different *types* of learning. But this fact is actually irrelevant to the question of a general factor in human intelligence, which we know to have a large genetic component and would therefore unquestionably respond to selection. Cronbach concludes this section by saying: "Jensen cites Scott as if he endorsed such an idea" [of a general learning ability in animals]. I did no such thing. As readers of my article can plainly see, I cited Scott & Fuller (*Genetics and the Social Behavior of the Dog,* 1965) along with Fuller & Thompson (1960) strictly in connection with my general introductory statement to this section, to the effect that behavioral traits respond to selective breeding in animal experiments. These are still the best two general references I can give for this statement.

Twin Studies

Kagan, a leading developmental psychologist, similarly criticizes parts of my paper in a way that hardly stands up under close examination. For example, he cites Gottesman, a behavioral geneticist, as questioning "the validity of Jensen's ideas." From Gottesman's article (1968, p. 28) Kagan reports: "In a study of 38

pairs of identical twins reared in *different environments,* the average difference in IQ for these identical twins was 14 points, and at least one quarter of the identical pairs of twins reared in different environments had differences in IQ scores *that were larger than 16 points.*" Gottesman, however, provided a bit more information. Actually *two* intelligence tests were used: a vocabulary test and a nonverbal test of abstract reasoning. The vocabulary test showed the average twin-pair difference of 14 points; the nonverbal test showed a difference of 10 points. Kagan himself italicized "*different environments,*" so let us look at the average difference on these tests between twins *reared together*: vocabulary = 9 IQ points, nonverbal = 9 IQ points. The average difference between the scores of the *same persons* tested twice on the same tests can be inferred from the reliabilities of these tests: vocabulary = 4 IQ points, nonverbal = 6 IQ points.[4] But the best way of seeing whether the Gottesman review cited by Kagan "questions the validity of Jensen's ideas" is to look at the original study which Gottesman summarized, which is one of the most careful and rigorous twin studies ever conducted (Shields, 1962). Shields' twin correlations are shown in Table 2. I ask, do these results "question the validity" of any of he statements in my article regarding the heritability of intelligence? To go on to say, as Kagan does, that the difference between members of identical twin pairs reared apart is larger than the average difference between black and white populations finds absolutely no support in this evidence! Kagan does not mention the statistical fact that the average absolute difference between twins includes the tests' measurement error, while the difference between the means of large groups does not contain this source of error.[4] The average absolute differences for height, intelligence, and scholastic achievement between a variety of kinships are shown in Figure 1.

In a similar vein of criticism is Hunt's comment: ". . . it is interesting to note what he [Jensen] omits from a paragraph quoted from the geneticist Dobzhansky," whom I quoted in part and paraphrased in part. Hunt's statement implies that the part of Dobzhansky I did not directly quote contradicts my own views. The omitted portion of Dobzhansky reads: "Although the genetically-guaranteed educability of our species makes most individuals trainable for most

[4] The standard error of measurement of most IQ tests is between 5 and 10 IQ points. This source of error is estimated by testing the same person twice or from split-half scores of odd vs. even numbered items.

TABLE 2

Correlations Between MZ Twins Reared Together and Apart[a]

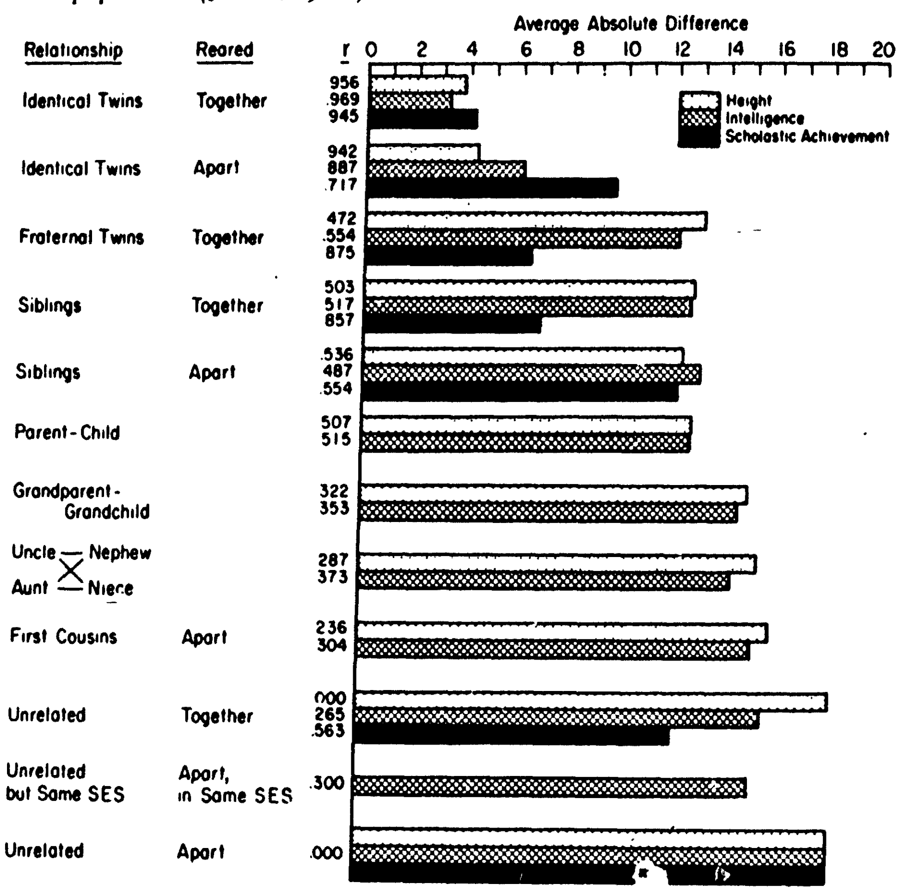
Measure	Twins Reared Apart ($N = 44$) r	Twins Reared Together ($N = 44$) r
Mill Hill Vocabulary	.74	.74
D48 Domino Test	.76	.71
Composite Intelligence Test Score*	.77	.76
Composite Intelligence Test Score Corrected for Attenuation	.86	.84
Height — Males	.82	.98
Height — Females	.82	.94
Weight — Males	.87	.79
Weight — Females	.87	.81
Extraversion**	.61	.42
Neuroticism**	.53	.38

[a] Source: Shields (1962), p. 69.
* Mill Hill Vocabulary Scale and the D48 (Domino) Test
** Maudsley Personality Inventory

occupations, it is highly probable that individuals have more genetic adaptability to some occupations than to others. Although almost everybody could become, if properly brought up and trained, a fairly competent farmer, or a craftsman of some sort, or a soldier, sailor, tradesman, teacher, or priest, certain ones would be more easily trainable to be soldiers and others to be teachers, for instance. It is even more probable that only a relatively few individuals would have the genetic wherewithal for certain highly specialized professions, such as musician, or singer, or poet, or high achievement in sports or wisdom or leadership." The reader can see for himself if Dobzhansky's statement in any way contradicts my own paraphrase.[5]

[5] The paraphrase read: "Some minimal level of ability is required for learning most skills. But while you can teach almost anyone to play chess, or the piano, or to conduct an orchestra, or to write prose, you cannot teach everyone to be a Capablanca, a Paderewski, a Toscanini, or a Bernard Shaw."

FIGURE 1.
Correlations, r, (corrected for attenuation, i.e., error of measurement) between persons with different degrees of kinship and reared together or apart. The average absolute difference (corrected for error of measurement) between pairs of individuals is based on the same scale for height, intelligence, and scholastic achievement, with a standard deviation (SD) of 16, the SD of Stanford-Binet IQ's in the normative population (Jensen, 1968a).



Individual Differences vs. Group Differences

Kagan further claims that my article contains "a pair of partially correct empirical generalizations wedded to a logically incorrect conclusion." The "partially cor-

rect" empirical generalizations he refers to are (a) the high heritability of intelligence (is there contrary evidence?) and (b) the average difference of about one standard deviation (15 or 16 IQ points) between Negro and white children on standardized intelligence tests (is there contrary evidence?). The "logically incorrect conclusion" is that, given these two facts, the IQ difference between Negro and white children, therefore, involves genetic as well as environmental factors. I have not drawn this "conclusion" from this premise, as geneticist Crow acknowledged in stating: "Strictly, as Jensen mentions, there is no carryover [of heritability measures] from within-population studies to between-population conclusions."[6] I have explained in greater detail elsewhere (Jensen, 1968b) that heritability coefficients by themselves cannot answer the question of genetic differences between groups, but when used along with additional information concerning the amount of relevant environmental variations within groups and overlap between groups, can enter into the formulation of testable hypotheses that could reduce the heredity-environment uncertainty concerning group differences. For example, we can pose the question: are differences (as measured by, say, median overlap) between various racial groups in the same society larger on mental tests of relatively low heritability than on tests of relatively high heritability within the groups being compared? Would not environmental and genetic hypotheses of the cause of the group difference lead to opposite predictions? Are these predictions operationally testable, just as other hypotheses in science? They have not, to my knowledge, been tested, and so, of course, I have not, contrary to Kagan's claim, drawn any conclusion about the outcome of such an hypothetical experiment. Also, other types of experiments permitting much stronger inference have been proposed but have not yet been done. I simply say there is sufficient evidence—and I present a list of items not mentioned by Kagan—to suggest it is not an unreasonable hypothesis that racial differences in mental abilities involve genetic

[6] Considered not as a test of genetic racial differences but merely as an abstract problem in quantitative genetics, I wonder if Crow would not agree with the following: Given two populations (1 and 2) whose means on a particular characteristic differ significantly by x amount, and given the heritability (H_1 and H_2) of the characteristic in each of the two populations, the *probability* that the two populations differ from one another genotypically as well as phenotypically is some monotonically increasing function of the magnitudes of H_1 and H_2. Such probabalistic statements are commonplace in all branches of science. It seems that only when we approach the question of genetic race differences do some geneticists talk as though only one or two probability values is possible, either 0 or 1. Scientific advancement in any field would be in a sorry state if this restriction were a universal rule. Would Crow argue, for example, that there is no difference in the *probability* that two groups differ genetically where H for the trait in question is .90 in each group as against the case where H is .10? In the absence of absolute certainty, are not probabalistic answers still preferable to complete ignorance?

as well as environmental and cultural factors. What factual or theoretical genetic evidence can Kagan present that this hypothesis is unreasonable or has already been scientifically rejected? Does Kagan advocate the fallacy that until a reasonable hypothesis has been definitely proved, we must believe that the *opposite* of the hypothesis is true? Or does he believe that these questions should not even be asked, much less formulated into testable hypotheses? My position is that reasonable hypotheses concerning socially and educationally relevant questions should be subjected to appropriate investigation and the findings be published and widely discussed by the scientific community and the general public as well.

The Bloom Fallacy

Cronbach notes that I refer to Benjamin Bloom's (1964) summary of age-to-age correlations of mental test scores up to 17 or 18 years of age. Cronbach believes that since I introduced this source I was also obligated to disclose that Bloom gives these data an interpretation opposite to mine. "Bloom sees the gains from year to year in test score as random and unpredictable, hence due to external events and not inheritance." I have no argument with Bloom's correlations, which are empirical fact. His interpretation of them, however, is fallacious, and though it does fit the correlation data themselves, it does *not* fit other data that are an essential part of the picture. These correlations, beginning at around zero between ages 1 and 18 years, gradually increase up to about .90 between ages 16 and 18. This pattern of correlations would result between series of scores if a number of random increments were added to each score starting with a base of zero (or some value without variance). But differences among the final scores, each consisting of the summation of random increments, will not be at all predictable. Yet we know that mental test scores are quite predictable, just from a knowledge of the parents' IQ's, even before the child is born. (The correlation of midparent and offspring at age 18 is about .70.) What the evidence on the heritability (*H*) of IQ tells us is that about 80 per cent of the variance in IQ's is conditioned by the genes, in other words, by factors already present at conception. This being the case, the interpretation of mental growth from birth to 18 years of age as a process of adding random increments just makes no sense. The Bloom model would be in accord both with the facts of the age-to-age correlations and with the facts of the heritability of IQ if it conceived of the adult level of ability as a genetically predicted level of ability from which random increments are subtracted, going in the backward direction toward birth. In other words, the genetic factors laid down at conception

77-942 O - 72 - 44

are increasingly realized in the individual's performance as he approaches the asymptote of that performance, in this case, ability on mental tests.

Cronbach also mentions late blooming in IQ, i.e., the fact that some persons show marked spurts in their relative position even as late as adolescence. Why should it be assumed that these mental growth spurts are environmentally caused? In fact, the relatively high correlation between identical twins across the whole age range, even in the range of the lowest year-to-year correlations, is a strong indication that genetic factors play a major part in the *form* of the individual's growth curve for intelligence, just as is true for height.

Underplaying the Role of Heredity

Cronbach says: "Jensen accuses writers on education of underplaying or denying the role of heredity. Some of this bias does exist, but Jensen is unfair. He does not quote the writers in psychology and education who do devote space to heredity." On the contrary, these are the ones about whom I have the greatest complaint. I do not criticize textbook writers who merely omit discussion of the heredity-environment issue. I *do* object to those textbook authors (Cronbach is *not* among them) who bring up the subject but then distort, misrepresent, or minimize the relevant evidence. I have recently surveyed 25 of the most widely used recent textbooks in educational psychology with reference to this topic and I am preparing a separate article on their treatment of the heredity-environment aspects of individual and group differences. Leaving out those few that say nothing about these topics, all but a few of the rest give what must be regarded as inaccurate or misleading information.

The Interval Scale of IQ

My argument that IQ's are approximately normally distributed in the population and that the IQ scale behaves like an *interval scale* is claimed by Hunt to be circular. Hunt shows that he misses the essential point when he says ". . . apparently, for Jensen, going twice around the circular argument removes its circularity?" The argument:

(a) We *postulate* that intelligence is normally distributed in the population, just as most other metrical biological characteristics (e.g., height, age of menarche, head circumference, etc.).

(b) We devise an intelligence test to yield a normal distribution of scores in a representative sample of the population. If intelligence is *in fact* normally distributed, and if our test scores yield a normal distribution, it necessarily follows

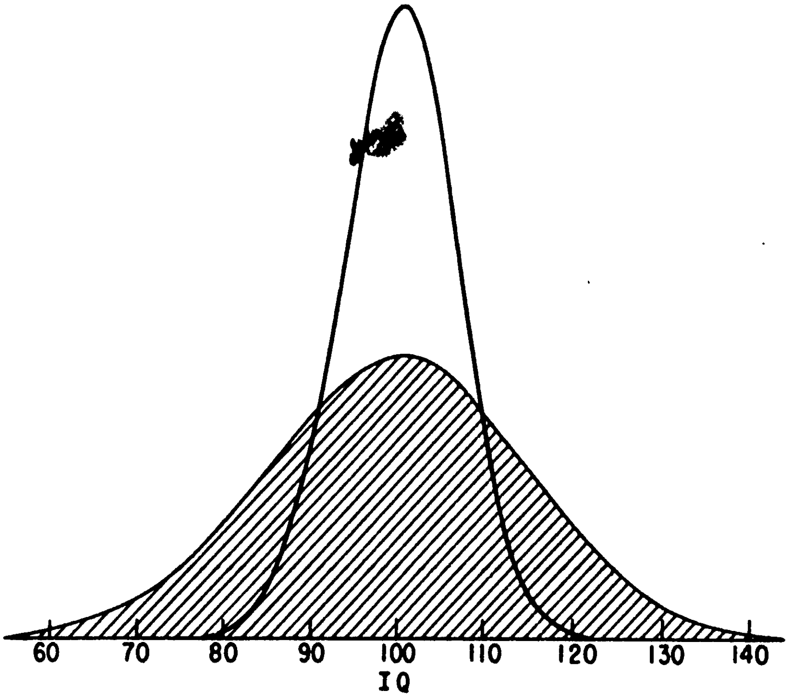
that the test scores constitute an equal interval scale. (If the scale were transformed, as by taking the square, square-root, logarithm, or any other non-linear transformation of the scores, the distribution would no longer be normal.) So far the logic is, of course, circular, as is the first step in *all* forms of measurement in science.

(c) But then we go beyond the circularity by determining if our postulate (i.e., normality) and the system of measurement that is relevant to it (i.e., interval scale) can make quantitative predictions of some phenomenon which is itself entirely independent of our assumption about the scale of measurement. If the prediction is then borne out in fact, the circularity is broken. The independent phenomenon we wish to predict in this present case is the regression of IQ for different degrees of kinship. The amount of regression for quantitative traits for various degrees of kinship is predicted from principles of population genetics and holds for clearly inherited metrical physical characteristics which are definitely known to be measured on an interval scale (e.g., height)—and our method of measuring intelligence itself plays no part in these genetic principles or analogous physical traits, so we are no longer involved in a circular argument. The genetic predictions will be borne out, however, only if our measurements of intelligence constitute an interval scale, because the genetic predictions assume rectilinear regression lines between kinship for metrical traits. The fact that the obtained regression lines for IQ's are rectilinear and closely in accord with the predictions (the same predictions that would be made for height, head circumference, fingerprint ridges, etc.) means that the IQ measurements behave like an interval scale. The genetic evidence, reviewed in my paper, fully supports this. Make a nonlinear transformation of the IQ scale and what happens? The kinship regressions are then clearly not rectilinear and the obtained kinship correlations are not in accord with the genetically predicted values. Furthermore, there is nothing in this whole argument which suggests, as Hunt accuses me of implying, that the present IQ distribution "is fixed in human nature for all time or until selective breeding alters it." Here Hunt again sets up his favorite straw man—"*fixed* intelligence."

The Editors' introductory summary of Hunt's paper says that "He [Hunt] finds Jensen's claims about the high heritability of intelligence unsubstantiated." Yet I find in Hunt's paper nothing that challenges either the theory or the methods or the findings concerning the numerous studies of the heritability of intelligence which are summarized in my article! If one wishes to argue with the empirical finding of a heritability coefficient (*H*) of, say, 80% for intelligence (the average value of *H* for the studies reported in the literature), then one must fault those

FIGURE 2.

Comparison of what the distribution of IQ's theoretically would be if all genotypes were identical (for IQ 100) in an "average" environment (assuming a normal distribution of environmental advantages) and all variance were due only to nongenetic (environmental) factors (heavy line). Under these conditions the heritability (H) of IQ's would be zero, instead of .80 as in the present population. The shaded curve represents the normal distribution of IQ's in the present population.

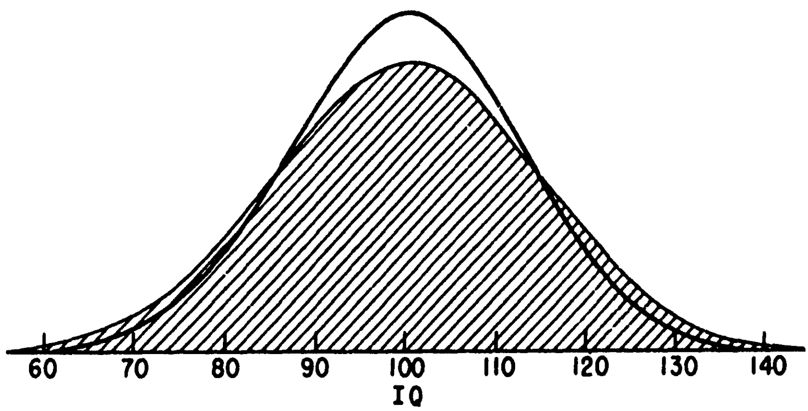


heritability studies which yield these results. Neither Hunt nor any of the other discussants has done this.

Phenotypic Variation of a Given Genotype

I wish to make it as clear as I know how just what a heritability (H) value of .80 actually means. Crow and Cronbach essentially reiterate what I said about the meaning of H. The latter says: "The index of .80 is impressive, but it is less discouraging than Jensen implies," and he presents a rather complex statistical argu-

FIGURE 3.
The theoretical distribution of IQ's if all variance due to environmental factors were eliminated (with everyone having an "average" environment) and all the remaining variance were due only to genetic factors (heavy line). Under these conditions the heritability (H) of IQ's would be 1.00. The shaded curve represents the normal distribution of IQ's in the present population, in which H = .80.



ment to indicate the range of phenotypic variation for a given single genotype which is implied by an H index of .80. The same argument can be illustrated perhaps more simply by graphical means. I did this in my original manuscript, but it was edited out, probably because it seemed redundant. But I think the graphical explanation is worth the space it takes. Figure 2 shows the normal distribution of IQ's in the population (shaded curve), and the heavy-line curve shows the hypothetical distribution of IQ's if all persons in the population had exactly the same genotype for intelligence and the only sources of variation were environmental. The area under both curves is the same, but the tall curve has only 20% of the variance (i.e., $1 - H = .20$) of the flat curve. In other words, it is the distribution of phenotypes for a particular genotype, given $H = .80$. This depicts essentially what Cronbach's statistical sortie was aimed to point out. But it is only *half* the picture. Figure 3 shows the reverse hypothetical situation, i.e., the difference in the IQ di tribution (heavy-line curve) if genotypes remained as varied as they actually are but everyone had the same environment (pre- and post-natal), which, of course, is possible only theoretically. The population vari-

ance in IQ's is thus reduced by 20%, and Figure 3 is how it would look. To point to only one or the other figure alone is improper. It takes *both* to tell the true picture.

Points of Misunderstanding

Confusion Between Population Average and Individual Differences

The most common point of confusion among several of the discussants concerns the distinction between common environmental factors that affect the population average and factors that account for individual deviations from the population average. Genetic and environmental factors are involved in both of these two aspects (i.e., population mean vs. individual variation), though not necessarily to the same degree. If the population average were not susceptible to environmental influences, there would, of course, be no value in education! Children can learn and do learn when appropriate opportunities are provided, just as they grow when food is provided. And the average level of developed skills in the population will reflect to an important degree the extent and quality of the opportunities for learning, just as the average stature of the population will reflect to some degree the quality of nutrition. While widespread improvement in the environment relevant to a particular trait may raise the mean level of the population on that trait, it does not necessarily, or even usually, decrease differences among individuals. No one denies the importance of certain environmental conditions for the development of phenotypic characteristics. What heritability studies of intelligence show, however, is that in the European and North American Caucasian populations in which these studies were conducted environmental variations account for relatively little (about 20%) of the variation in intelligence among individuals. These studies by themselves can tell us nothing about changes in the mean of the population across generations. Even though the offspring may be brighter or taller than their parents, the *correlation* between parents and children does not change appreciably. For highly heritable traits, like intelligence, parental phenotypes thus remain a statistically reliable basis for predicting the deviations of their offspring from the population mean. Improving the population's relevant environment for the development of a trait usually *increases* the phenotypic manifestations of genotypic differences, and, as Bereiter points out, it *increases* the heritability of the trait: "One's view of the future beyond equality of opportunity must, therefore, be of a future in which differences in intelligence are virtually one hundred percent determined by heredity." Bereiter adds in a footnote: "This

eventuality is in no wise to be forestalled by individualized instruction or any more libertarian tactic; on the contrary, such approaches should allow inherited differences to reach full flower, as advertised in the slogan, 'enabling each child to realize his fullest potential.'"

This brings us to the question of the primary aims of compensatory education. The aims are often explicitly stated as being to decrease or remove the scholastic (and ultimately occupational) achievement gap between children called disadvantaged and the rest of the population, or even to make all children perform at least at the population average for their grade level throughout their years in school. Educational innovations, improvements in instructional techniques, and so on, when they are successful, are just as likely to increase the learning and achievements of the advantaged as of the disadvantaged, with little if any decrease in the relative differences among individuals, so that E. L. Thorndike's dictum would remain valid: "In the actual race of life, which is not to get ahead but to get ahead of somebody, the chief determining factor is heredity." Equality of opportunity is a worthy and attainable goal. Equality of performance is a misguided hope. The important thing for the welfare of children and of society in general would seem to be to try and create conditions that will maximize the proportion of the population that can learn and work successfully and rewardingly in the diverse occupational roles that the society provides. It is clear that various peoples and societies in the past and in the present have approached this realistic goal to quite different degrees, and it would seem worthwhile to inquire into the social, biological, and educational conditions which have either hindered or promoted the realization of this goal. I would hypothesize that among the relevant conditions would be at least two prominent factors: (a) the working of eugenic pressures, either consciously and directly, or indirectly through the value system, social structure, socially-conditioned mating patterns, and the like, and (b) a wide diversity of educational options, paths, and goals.

Height as an Example

I have said that the mode of inheritance of intelligence quite closely parallels that for physical stature. Four of the discussants referred to the overall increase in height in the population as if this fact somehow diminished the importance of heredity in individual differences in height, and even more so in intelligence, since intelligence has a lower heritability than height (about .80 vs. .95). Because this has been one of the commonest arguments put forth by persons traditionally called environmentalists, I think it deserves a closer look than it was given by the discussants. The

parallel between height and intelligence is close enough that we may gain some insights about the latter from a study of the former, about which much more is known concerning population trends across many generations.

Crow states that because of unidentified environmental influences height has increased by a "spectacular amount." And Hunt, on the basis of what he heard from guides at Jamestown's Festival Park and aboard the *U. S. Constitution,* states that height "appears to have increased *nearly a foot* without benefit of selective breeding or natural selection." Presumably Hunt is referring to the increase in adult height since about the 17th century. The implication is that all of this increase in height is strictly the effect of environmental and not genetic factors.

Let us see what more dependable authorities than tourist guides have to say about this subject. I have obtained my information from a book on human genetics by a noted British geneticist (Carter, 1962), and from comprehensive articles on this subject by J. M. Tanner (1965, 1968), the world's leading researcher on human growth. Here is what I find:

First of all, it is essential to distinguish between growth *rate* and final (adult) *level.* Adult height has increased little over the past century or so. Carter (p. 102) says that skeletal remains suggest there has been little appreciable change in height in Britain over the past 5000 years. "If there has been any increase [in adult height in Britain] it is only of the order of 1 inch. What environmental improvements appear to be doing is, in the main, to accelerate growth, so that full adult height is being reached earlier. Records from the armed services, prisons, and anthropological surveys suggest that full adult height has not changed by more than 1-½ inches for the past century" (p. 102). Other countries have shown slightly higher increases than in Britain, and Tanner (1968) concludes that adult height has increased 2-½ to 3-½ inches in the past century. Increases before the last century were relatively minute. While the increase in height since about 1700 was a positively accelerated curve, it has become negatively accelerated in the 20th century, and the trend is leveling off, especially in the United States. Growth *rate,* and consequently children's height, has shown much greater increases. Children now attain their full adult height by 18 or 19, on the average, rather than at 26, as was the case only 50 years ago. The trend toward earlier maturation shows up most dramatically in the lower age of menarche, or first menstrual period, which has declined from 17 to 13 years of age since 1840.

The trend toward earlier maturity seems to be related largely to environmental factors—probably improved nutrition and, it has been hypothesized, electric lights. (Children today spend more time awake and, due to electric lighting, more hours

Reducing the Heredity—Environment Uncertainty
ARTHUR R. JENSEN

under illumination, so they grow for more hours per day, just as chickens raised under constant illumination reach egg-laying capacity much younger than when raised under normal conditions.) But part of the cause of increased growth rate is also genetic. The increase in *adult* height may be almost entirely attributable to genetic factors. Tanner (1965) points out that among environmental factors the increase in adult height is at least as closely related to the introduction of the bicycle and other improved modes of transportation as to improvements in nutrition and health care. What is the explanation? It is what geneticists call the outbreeding effect, heterosis or hybrid vigor. Tanner (1968) states that the "height of adults is significantly and inversely correlated with the degree of inbreeding in the region studied," and "the trend in adult height may have in whole or in part a genetic explanation." It has been estimated that 10 to 20 per cent of the variance in height is due to genetic dominance, so that the mean of the offspring of two parents will not be halfway between the parents but slightly closer to the taller parent. Outbreeding increases heterozygotes in the population with a consequent increase in height. This heterosis due to outbreeding also enhances growth rate and early maturation as amply demonstrated in numerous experiments in animal breeding. Outbreeding has increased at a steady rate ever since the introduction of the bicycle. For example, sons of parents who were from *different* Swiss villages were taller by approximately 1 inch than the sons of parents from the *same* village. Persons born to parents whose inbreeding is to the degree of first cousins average 1.4 inches shorter than persons whose parents are unrelated. According to Tanner, the average degree of outbreeding that has taken place in the last century can account for 0.8 inches increase in height per generation. The increase in heterozygosity, of course, eventually "saturates," and the effects level off, as has already occurred in the U. S. That genetic as well as nutritional factors are a major cause of the increase in actual height is further shown in the fact that approximately the same increase has occurred in all social classes in Western countries even though there have been nutritional differences among social classes. On the other hand, earlier maturation, as indexed by age of menarche, is more related to nutrition, as shown by a decrease in social class differences in countries with a very wide range of nutrition. Thus Hong Kong has shown a convergence between social classes in the decreasing age of menarche, while England and Scotland have not.

Have genetic differences between individuals and between groups *decreased* with the average increase in height in the population? No. Take the sex difference in height, which is surely genetic. Since males have responded more than females to improved nutrition, the sex difference in height has slightly increased. The range

of individual differences in height is at least as great as ever it was and the heritability of height is probably higher than it has ever been.

Thus the slight increase in the population's mean height over the last two centuries—the environmentalists' favorite counter-argument to the high heritability of IQ—itself turns out to be largely a genetic phenomenon!

What has been said about height probably applies also to intelligence and other biologically-conditioned characteristics. There is some evidence, for example, of an increase in intelligence test performance in the general population between World War I and World War II (Tuddenham, 1948), due no doubt to improvements in education, nutrition and health care, and standards of living in general, and the same general factors involved in the increase in height. Intelligence variance, too, has a genetic dominance component not very different from height. Both white and Negro populations have shown the reported increase in intelligence test performance, but there has been no indication of a *convergence* of their mean scores since World War I, although there have been marked socioeconomic and educational advances since then. In fact, there is some indication from armed forces tests and nationwide testing surveys that, if anything, the average difference in performance between Negro and whites may have *increased* since World War I (e.g., Minor, 1957).[7]

Confusion of Cultural Disadvantage with Sensory Deprivation

Hunt's paper places great emphasis on the role of sensory stimulation in early development as a factor in later mental attainments. He cites particularly two classes of evidence in support of this hypothesis: (a) experiments on the effects of extreme sensory deprivation in animals, and (b) observations of children subjected in early infancy to extreme sensory deprivation and motor restriction through being confined in cribs in understaffed orphanages.

The connection between these lines of evidence and the average lower IQ's and deficiencies in scholastic performance of children called culturally disadvantaged is purely hypothetical. I seriously question the relevance of these types of evidence for understanding the observable abilities of disadvantaged children.

I do not contest the evidence showing that rabbits, kittens, and chimpanzees

[7] It has also been argued that our concern should be with the relative improvement of Negroes compared to the white population, rather than with the absolute improvement of one group. Though some differentials have been cut, a time-gap analysis indicates that the Negro lags about a quarter century behind the white, and this lag has not been reduced since World War I. On some measures, there is evidence that the environmental differences, expressed in time-lag, are increasing. See: Rashi Fein, "An Economic and Social Profile of the Negro American," *Daedalus*, 94, no. 4 (Fall, 1965), 815-846.

after being reared in total darkness manifest irreversible histological effects, such as degeneration of the optic disc, optic nerve, pyramidal cells in the striate area, and so on. Culturally disadvantaged children are obviously not reared in the dark. The experiments cited by Hunt are interesting but irrelevant to the problems discussed in my paper. Somewhat more relevant are Harlow's experiments (cited in my paper) on primates reared under severe sensory-motor deprivation but not the absence of light which results in optic-neural degeneration. Harlow's deprived monkeys were reared in isolation in small, lighted cages with uniform opaque walls and containing few manipulanda. Yet after prolonged periods of being raised in such an environment they showed no deficiencies in learning performance as compared with monkeys raised together in large, open cages permitting a variety of sensorimotor experience. Similar sensory deprivation and enrichment studies using rats, such as the work of Krech and Rosenzweig cited by Hunt, are clearly less relevant than the primate experiments in their implications for human behavior. It should be noted, however, that even in the case of rats, the greatest extremes of rat environment, from deprived to enriched (where the enrichment includes experience in mazes), that have been devised in the laboratory result in differences in maze learning ability only about one-fourth as large as those produced genetically by selective breeding for maze learning.

Hunt also attaches importance to experience in the development of sensorimotor integration, referring to experiments with rats climbing a guy-rope, which suggests that "each coordination, between vision-and-hand motion or between eye-function and ear-function, has its own neuro-electrical-chemical-anatomical equipment . . . When such equipment has emerged as the consequence of a given bit of functional accommodation or learning, it can readily be employed in other functioning and thereby becomes the basis for the transfer of training." But do such elemental components of sensorimotor accommodation and integration have any less chance to develop in a slum than in a penthouse? It seems far-fetched to me that, as Hunt suggests, these components of early sensorimotor development form the basis of Spearman's *g* or general intelligence factor. I cited evidence in my paper showing that, if anything, there is either a zero or a negative correlation between most indices of early behavioral development, such as the Bayley Infant Scales, and later IQ. Kagan (1966) has identified some components of early behavior which apparently show a more marked correlation with later intelligence than is generally found in the standard infant scales of development. Kagan reports that on certain laboratory tests of cognitive functioning lower-class children, as early as 8 to 12 months of age, show slower rates of information processing than middle-class children of the same ordinal position among their siblings. Kagan observes:

Lower-class children show less rapid habituation, less clear differentiation among visual stimuli, and, in a play situation, show a high threshold for satiation. The latter measure is obtained by placing the child in a standard playroom with a standard set of toys (quoits on a shaft, blocks, pail, mallet, peg board, toy lawn mower, and toy animals) and by noting the time involved in each activity. Some children play with the blocks for 10 seconds and then skip to the quoits or the lawn mower, playing only 10-20 seconds with each individual activity before shifting to another. A second group of children, called "high threshold for satiation infants" spends 1 or 2 minutes with an activity without interruption before changing. We do not believe the latter group of infants is taking more from the activity; rather it seems that they are taking longer to satiate on this action. It is important to note that the observation that lower-class infants show a high threshold for satiation contrasts sharply with the observation that 4-year-old lower-class children are distractible and hyperkinetic. We believe both descriptions. The paradox to be explained is why these lower-class children are pokey and lethargic and nondistractible at 12 months of age, yet display polar-opposite behaviors at 48 months of age (Kagan, 1966, pp. 105-106).

The other line of evidence appealed to by Hunt is on orphanage infants deprived of normal sensorimotor experience during the first one to two years of life, as in the well-known study by Skeels and Dye (1939). After such deprivation, these children have very retarded developmental quotients and their entire behavior is in marked contrast to that of children typically called disadvantaged. After placement in good environments, the children showed an average gain of about 30 IQ points, became average children, and grew up to be average adults (Skeels, 1966). This, too, is in contrast to typical disadvantaged children, who, rather than showing a tendency to catch up when placed in a presumably more culturally enriched environment—the school—begin gradually to fall behind in cognitive development. The typical characteristics of culturally disadvantaged children are a different set of phenomena from those resulting from early sensory deprivation. The contrast is further highlighted by studies of children who suffer severe verbal deprivation as a result of being born completely deaf: These children show a very marked retardation, usually amounting to one to two years, on tests of verbal intelligence. Unlike disadvantaged children, however, the deaf children, despite continuing deafness, gradually catch up in intellectual performance—it merely takes them longer to acquire information because of their severe sensory handicap. But once acquired, normal mental development continues. In one of the most careful studies of mental development in deaf children, the authors concluded that the deaf merely take *longer* to reach the same level of verbal-conceptual-thinking ability as normal persons. The authors state: ". . . the differences found between deaf and hearing adolescents were amenable to the effects of age and education and were no longer

Reducing the Heredity—Environment Uncertainty
ARTHUR R. JENSEN

found between deaf and hearing adults. Dissociation between words and referents, verbalization adequacy, and level of verbalization were not different for deaf and hearing subjects. Our experiments, then, have shown few differences between deaf and hearing subjects. Those found were shown to fall along a normal developmental line and were amenable to the effects of increased age and experience, and education" (Kates, Kates, & Michael, 1962, pp. 31-32).

How much of Hunt's association of sensory deprivation with the culturally disadvantaged has affected psychologists' perceptions and descriptions of the environment of infants of mothers called culturally disadvantaged? Note Kagan's description of children he has studied in the lower-class white population: "... the lower class mothers spend less time in face to face mutual vocalization and smiling with their infants; they do not reward the child's maturational progress, and they do not enter into long periods of play with the child. Our theory of mental development suggests that specific absence of these experiences will retard mental growth and will lead to lower intelligence test scores." There is not unanimous agreement that the culturally disadvantaged have such impoverished interpersonal interactions in infancy as described by Kagan. The early environment of Negro infants, for example, is described in quite contrasting terms by a Negro writer, Kristin Hunter: "Ghetto babies must be the most thoroughly loved in the world; they are passed from loving arms to loving arms, cradled, cuddled, tickled, endlessly discussed and admired" (Hunter, 1969). This does not sound like sensory deprivation.

In emphasizing the environments of the extreme poor, Hunt remarks that "few if any of the studies of heritability have included the truly poor, so they have missed this portion of the variation in the circumstances of rearing." Heritability studies have included all social classes, but I agree that special attention should be given to including the very extremes of the existing environmental continuum. One might also expect, however, that sampling from an increased range of environments will simultaneously yield a correlated increase in genetic variation, thereby leaving the heritability of IQ approximately the same.

Hunt also seems to assume that anything that will accelerate any aspect of development is psychologically good and will have enhancing effects on later mental ability. This is sheer speculation without empirical support. Putting mobiles over a child's crib may very well bring about an earlier eye-blink response in infants, but what has this to do with the mental abilities measured by IQ tests and correlated with scholastic performance? There is just no evidence that these types of stimulation in early infancy, over and above what infants normally get, are in any way

related to their intelligence at school age. In fact, there is some evidence, again from primate experiments, that attempting to develop abilities ahead of the normal maturation of cognitive processes may even be harmful. Harlow (1959), for example, found that very young monkeys have much greater difficulty than somewhat older monkeys in learning-set formation (i.e., "learning to learn") but that the younger monkeys can acquire learning sets by being given much more training than is needed by older monkeys. The younger monkeys, however, do not attain the same level of proficiency in these problems. The more important fact is that the younger monkeys cannot be trained to do as well as the older monkeys even when they finally reach the same age as the monkeys who trained at a later age. Harlow states: "... these data suggest that the capacity of the two younger groups to form discrimination learning sets may have been impaired by their early, intensive learning-set training, initiated before they possessed any effective learning-set capability." The more advanced cognitive structures awaiting later brain maturation apparently were never invoked in the earlier trained monkeys, whose performance remained permanently below that of monkeys trained at a later age. This observation would seem to be consistent with Elkind's conjecture that "... the longer we delay formal instruction, up to certain limits, the greater the period of plasticity and the higher the ultimate level of achievement."

Associative and Cognitive Abilities

My theory of two broad categories or clusters of mental abilities, labeled Level I and Level II because they seem to stand in some hierarchical relationship, is somewhat misinterpreted by Cronbach and Hunt. In factor analyses, a variety of tests of associative learning ability and memory (digit span, serial and paired-associate learning, free recall of uncategorized lists, etc.) tend to cluster together; these tests represent in varying degrees what I call Level I abilities. On the other hand, another class of tests, which are not highly correlated with Level I tests also cluster together: standard verbal and nonverbal IQ tests, tests involving abstract reasoning, symbol manipulation, free recall of conceptually categorized lists, etc. I call these abilities Level II.

Hunt lists a great variety of types of learning associated with various experimental techniques for the laboratory study of learning identified with Ebbinghaus, Pavlov, Thorndike, Hull, Skinner, and Piaget, and then says that my broad distinction between associative and cognitive learning is "but a conceptual drop in the bucket." This is to miss the point that Level I and Level II represent broad categories of abilities which do emerge in factor analyses, and many of the types of learning listed

by Hunt can be represented in this two-dimensional factor space. The fact that one can fractionate these broad factors does not detract from their scientific usefulness in attempting to understand the structure of mental abilities. Nor is it meaningful to call this theory an "over-simplification" as does Cronbach. It *is* a simplification of a diversity of phenomena, to be sure, but an essential aim of science is to conceptually organize and simplify disparate and variegated phenomena. There is no doubt of the complexity inherent in my formulation. For example, few, if any, tests can be regarded as measuring purely Level I or Level II under all conditions. We already know that paired-associate learning tests can be either Level I or Level II, or any admixture of the two, depending upon a number of experimentally manipulable variables. For instance, if the subjects (college students) are forced to learn a list of paired-associates at a very fast rate of presentation, the test, when included in a factor analysis, is loaded almost entirely on the Level I factor. If the same paired-associates are presented at a much slower rate, the learning scores are then substantially loaded on the Level II factor. Also, certain instructional techniques may change what are usually perceived as rote-learned tasks into conceptually mediated learning. Cronbach should be assured that I recognize a *continuum* of the susceptibility of various tasks to manipulation with respect to their Level I-Level II loadings. Some tasks are relatively easy to manipulate in this respect—for example, paired-associate learning and probably free recall of clusterable lists. Other tasks are much more difficult to manipulate through instruction, for example, the ability of children under 6 or 7 to copy the figure of a diamond, or to conserve volume in the Piagetian paradigm.

All this does not mean, however, that stable individual differences in Level I and Level II abilities do not exist or are trivial. Cronbach points out that spatial ability, which is highly heritable, can be improved through training. I hope he does not believe that this implies that the training will wipe out, or even decrease, individual differences in spatial ability or will lower its heritability within the group that received the training. There is good reason to believe that just the opposite would occur. I have found in some of my own research, for example, that prolonged practice (by college students) on digit span tests significantly *increases* the amount of reliable variance due to individual differences. All subjects improve with practice, but reliable individual differences become accentuated at the asymptote of improvement. Cronbach knows that when we talk about the heritability of an ability, we are not referring to the absolute level of performance that can be attained, but to individual differences in performance and the proportion of their variance attributable to genetic differences.

The Hope of the Instruction × Individual Differences Interaction

Hunt and especially Cronbach share the same hope I expressed in my paper (and on numerous other occasions) that the improvement of scholastic achievement and the minimization of individual and group differences in performance may be brought about by making use of the idea of a subjects × instruction interaction. In the simplest terms this means that if Jim and Bill are taught in the same way, they will differ more in how fast and how much they learn than they would if each one were taught by a different method which is especially suited to each child's individual pattern of abilities. Bereiter is clearly much less optimistic than the rest of us about the practical possibilities implied by the instructional interaction notion. His cogent remarks have indeed had a somewhat sobering effect on my own thinking on this topic and I have gone back to the literature to see how much hard evidence I could find to bolster my hope that this interaction notion of more individualized instruction holds the promise of solving our major educational problems. To my dismay, but in all fairness to Bereiter, I must admit that I can find very little evidence of pupil × type of instruction interaction in the realm of learning school subjects or for complex learning in general. Most of the evidence for such pupil × instruction interactions has been reviewed by Cronbach (1967) in a paper which is a "must" in this field. I believe that research based on a more fine-grained approach to the analysis and manipulation of instruction will be necessary before we can properly assess the educational potential of the pupil × instruction interaction. We do know that quite clear-cut interactions have been shown in laboratory experiments on simple learning tasks in which the tasks and methods themselves impose great constraints on what the subject can do in the learning situation. Then we can find significant interactions between learners and experimental variables (Jensen, 1967). When tasks are complex, involving a variety of abilities, as in school learning, and when there are few constraints on how subjects can learn, pupil × instruction interactions either fail to appear or are undetectable. At this point, indeed, I can only say it is my conjecture, my hope, that the Level I-Level II distinction may interact with instructional techniques to decrease the spread between disadvantaged and advantaged children in their mastery of the basic scholastic skills. I hope a variety of research will be directed to testing this hypothesis.

Cronbach solves no problem by saying "Capability is not at issue when a child does not call upon an ability he possesses." What about the ability to call up relevant subabilities and past learning when confronted with a new problem? This ability to transfer learning from one type of problem to another is the essence of intelligence; it is a Level II process. Why does the 5-year-old fail to copy a diamond

despite his ability to draw straight lines? Why does a child who has learned to add, subtract, multiply, and divide often fail in arithmetic "thought problems" which call upon the applications of these subabilities? It is the appropriate calling up, integration, and transfer of various subskills that constitute what we mean by intellectual capability. I can play chess; I know all the moves. But why can't I play like Alekhine or Capablanca? Is it simply because I do not call upon an ability that I possess? I doubt it.

Bereiter is correct, I believe, in his argument that complex intellectual tools act as amplifiers rather than equalizers of basic differences in problem-solving ability. Cronbach's argument that the invention of the computer has increased man's mathematical capacity has as much to do with individual differences in mathematical ability as the invention of the automobile has to do with individual differences in running ability.

Genetic Social-Class Differences in Intelligence

Because of differences between child-rearing practices of the middle-class and those of people of poverty, Hunt doubts that socioeconomic status (SES) differences in intelligence have any genetic component. If Hunt's supposition were true that there is no genetic component to social class intelligence differences, it would have to mean that all the factors involved in social mobility, educational attainments, and the selection of persons into various occupations have managed scrupulously to screen out all variance associated with genetic factors among individuals in various occupational strata. The possibility that the selection processes lead to there being only environmental variance in intelligence among various socioeconomic groups and occupations—a result that could probably not be accomplished even by making an explicit effort toward this goal—is so unlikely that the argument amounts to a *reductio ad absurdum*. If individual differences in intelligence are due largely to genetic factors, then it is virtually impossible that average intelligence differences between social classes (based on educational and occupational criteria) do not include a genetic component.

The argument is as follows: Twin studies and other methods for estimating the heritability of intelligence have yielded heritability values for the most part in the range from .70 to .90, with a mean value of about .80. Heritability (*H*) indicates the proportion of variance in a metric characteristic, such as height or intelligence, that is attributable to genetic factors. (Since the heritability estimate is derived from studies in European and North American Caucasian populations, the present genetic analysis of SES differences cannot be generalized across racial groups.)

$1 - H = E$, the proportion of variance due to non-genetic or environmental factors, which of course include prenatal as well as postnatal influences. The correlation between phenotypes (the measurable characteristic) and genotypes (the genetic basis of the phenotype) is the square root of the heritability, i.e., $\sqrt{H}$. An average estimate of $\sqrt{H}$ for intelligence is .90, which is the average correlation between genotype and phenotype. An estimate of the average correlation between occupational status and IQ is .50. What Hunt is saying, essentially, is that the correlation between IQ and occupation (or SES) is due entirely to the environmental component of IQ variance. In other words, this hypothesis requires that the correlation between genotypes and SES be zero. So we have correlations between three sets of variables: (a) between phenotype and genotype, $r_{pg} = .90$; (b) between phenotype and status, $r_{ps} = .50$; and (c) the hypothesized correlation between genotype and status, $r_{gs} = 0$. The first two correlations (r_{pg} and r_{ps}) are determined empirically and are represented here by average values reported in the literature. The third correlation (r_{gs}) is hypothsized to be zero by those who believe genetic factors play a part in *individual* differences but not in SES *group* differences. The question then becomes: is this set of correlations possible? The first two correlations we know are possible because they are empirically obtained values. The correlation seriously in question is the hypothesized $r_{gs} = 0$. We know that mathematically the true correlations among a set of variables, 1, 2, 3, must meet the following requirement:

$$r^2_{12} + r^2_{13} + r^2_{23} - 2r_{12}r_{13}r_{23} < 1$$

The fact is that when the values of $r_{pg} = .90$, $r_{ps} = .50$ and $r_{gs} = 0$ are inserted into the above formula, it yields a value greater than 1.00. This means that r_{gs} must in fact be greater than zero.

Another way of regarding this problem is as follows: If only the E (environmental) component determined IQ differences between status groups, then the H component of IQ's would be regarded as random variation with respect to status. Thus, in correlating IQ with status, the IQ test in effect would be like a test with a reliability of $1 - H = 1 - .80 = .20$. That is to say, only the E component (.20) of the total variance is not random with respect to indices of SES. Therefore the theoretical maximum correlation that IQ could have with SES would be close to $\sqrt{.20} = .45$. This value is slightly below but very close to the average value of obtained correlations between IQ and SES. So if we admit no genetic component in SES differences, we are logically forced to conclude that persons have been fitted to their socioeconomic status (meaning largely educational attainments and occupa-

tional status) almost *perfectly* in terms of their environmental advantages and disadvantages. In other words, it would have to be concluded that persons' innate abilities, talents, and proclivities play no part in educational and occupational selection and placement. This seems a most untenable conclusion. The only way one can logically reject the alternative conclusion, that there are average genetic intelligence differences among SES groups, is to reject the evidence on the heritability of individual differences in intelligence. But the evidence for a substantial genetic component in intellectual differences is among the most consistent and firmly established research findings in the fields of psychology and genetics.

Social and Educational Policy and the Heritability of Individual Differences

Cronbach states it is regrettable that I do not spell out the policies that should follow from my formulations and conclusions. This is, of course, another job. I am not a social or educational philosopher and I am sure that neither I nor anyone else at present has thought through all the policy implications of my article. I do believe that educational policy decisions should be based on evidence and the results of continuing research—and not just the evidence which is comfortable to some particular ideological position, but *all* relevant evidence. I submit that the research on the inheritance of mental abilities *is* relevant to understanding educational problems and formulating educational policies. For one thing, it means that we take individual differences more seriously than regarding them as superficial, easily-changed manifestations of environmental differences. And it means we look more critically and carefully at environmental variables that contribute most to differences in mental development, as I suggested that prenatal and nutritional factors had not been given due consideration. Also, it means we expend more research effort on exploring and mapping a wider range of abilities than those measured by IQ tests, on discovering the particular learning strengths of each child, and on devising methods that will more fully utilize these strengths to help all children to benefit more from their schooling. To refrain from discussing some of the relevant factors that should be considered in formulating policy simply because the details of such policy cannot yet be spelled out is, in my opinion, practically equivalent to saying: "Don't ask any questions unless you already know all the answers."

Brazziel's letter seems to be saying in part that my paper should not have been published in the first place. I would plead for more faith in the wisdom of the First Amendment. To refrain from publishing discussions of research on socially important issues because possibly there will be some readers with whose interpretation

or use of the material we may disagree is, in effect, to give those persons the power of censorship over the publication of our own questions, findings, and interpretations. It is only when all the available facts, issues, and questions can be openly examined and discussed by everyone that we can put any stock in the maxim that "the truth will out." I resent Brazziel's statement that I expound a theory of white supremacy, but I suppose it must be evaluated in the context of his overall reaction to my article. On this point, however, it might be of interest to some to note that on the basis of the evidence I have been able to review so far, if I were asked to hypothesize about race differences in what we call *g* or abstract reasoning ability, I would be inclined to rate Caucasians on the whole somewhat below Orientals, at least those in the United States. A case can be made for this conjecture on the basis of existing evidence, but this is not the appropriate place for it.

Reducing the Uncertainties

One disappointment with the discussions of my paper is the fact that attitudes of "let's not talk about genetics," or "it's too complicated," or "we can't find out the answers anyway," and so on, have prevailed over the attitude of inquiry and the application of intellectual ingenuity in trying to reduce our heredity-environment uncertainty. If there are weaknesses in the methods and the evidence I have presented, and of course there inevitably are at this stage, we would do well to note them as a basis for seeking more refined research methods and more and better data, rather than as a basis for minimizing the scientific and social importance of these questions, or sweeping them under the rug.

Brazziel is quite correct in noting, for example, that the Negro population of the United States, like the white, is very far from being genetically or racially homogeneous. In fact, it is doubtful that any babies of pure African descent are being born in the United States today, unless they are born to African exchange students. But Africans, too, are genetically heterogeneous. A number of studies based on the differential frequencies of various blood groups in African and Caucasian populations have shown that, on the average, persons socially classified as American Negroes now have an admixture of 20 to 30 per cent Caucasian genes (Reed, 1969). The percentage of Caucasian admixture varies greatly in various regions of the country, going from an average of below 10% in some Southern states to above 25% in some Northern states. These figures can be estimated with considerable precision in large population samples, depending on the number of different blood groups and other genetic polymorphisms one is able to take into account. With these methods individuals, too, can be categorized by proportions

of Negro-Caucasian admixture on a probabalistic basis. Possibly these same genetical techniques could provide a basis for more refined and accurate tests of hypotheses concerning racial differences in ability patterns. Since skin color is but poorly correlated with the percentage of Caucasian admixture, and because it may have social-environmental consequences, it could be statistically controlled in studies of the correlation between Negro-Caucasian admixture and measures of psychological characteristics. Environmental differences would not be an obstacle, since there is a wide range of racial admixtures in any large sample from highly similar environments. In fact, where there are half-siblings, intra-family comparisons might be possible, thereby controlling a host of environmental family-background factors. Other quite different approaches are possible, or a number of methods used in combination. The finding that electroencephalographic visually-evoked potentials are related to IQ means that intelligence might be measured on a physiological level, and such a measure would come closer than anything we now have to a true culture-free test. Studies of foster children of one race or social class adopted by parents of another is one more avenue. Such are only a few of the possible suggestions. Geneticists should be able to evaluate these and come up with better ideas. Collaborative research by geneticists and behavioral scientists could surely advance our scientific knowledge of racial and social class differences. To argue to the contrary, it seems to me, is to claim the impotence of a scientific approach and of human ingenuity, an attitude which is clearly contradicted by our great advances in other fields of inquiry. If the heredity-environment uncertainty is unresolvable in the sense that, say, perpetual motion is impossible, we should at least not be satisfied until we have discovered precisely the laws of nature which make it so.

It is already apparent that my article "How Much Can We Boost IQ and Scholastic Achievement?" has been eminently successful in widely provoking serious thought and discussion among leaders in genetics, psychology, and education concerning important fundamental issues and their implications for education. I expect now that this will stimulate further relevant research as well as efforts to apply the knowledge gained thereby to educationally and socially beneficial purposes. The whole society will benefit most if scientists and educators treat these problems in the spirit of scientific inquiry rather than as a battle field upon which one or another preordained ideology may seemingly triumph.

References

Bloom, B. S. *Stability and change in human characteristics.* New York: Wiley, 1964.

Carter, C. O. Differential fertility by intelligence. In J. E. Meade & A. S. Parkes (Eds.), *Genetic and environmental factors in human ability*. New York: Plenum Press, 1966. Pp. 185-200.

Cronbach, L. J. How can instruction be adapted to individual differences? In Gagné, R. M. (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967. Pp. 23-39.

De Lemos, M. Murray. The development of the concept of conservation in Australian aboriginal children. Unpublished Ph.D. dissertation. University of Western Australia, Nov., 1966.

Fuller, J. L., and Thompson, W. R. *Behavior genetics*. New York: Wiley, 1960.

Gottesman, I. I. Biogenetics of race and class. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.), *Social class, race, and psychological development*. New York: Holt, Rinehart & Winston, 1968. Pp. 11-51.

Harlow, H. F. The development of learning in the Rhesus monkey. *Amer. Sci.*, 1959, **47**, 459-479.

Hunter, Kristin. Pray for Barbara's baby. *Philadelphia Magazine*, Aug., 1968. (Reprinted in *Reader's Digest*, January, 1969).

Jensen, A. R. Varieties of individual differences in learning. In Gagné, R. M. (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967. Pp. 117-135.

Jensen, A. R. Social class, race, and genetics: Implications for education. *Amer. Educ. Res. J.*, 1968, **5**, 1-42. (a)

Jensen, A. R. Another look at culture-fair tests. In *Western Regional Conference on Testing Problems, Proceedings for 1968*, "Measurement for Educational Planning." Berkeley, Calif.: Educational Testing Service, Western Office, 1968. Pp. 50-104. (b)

Kagan, J. A developmental approach to conceptual growth. In H. J. Klausmeier & C. W. Harris (Eds.), *Analyses of concept learning*. New York: Academic Press, 1966. Pp. 95-115.

Kates, Solis L., Kates, W. W., & Michael, J. Cognitive processes in deaf and hearing adolescents and adults. *Psychol. Monogr.*, 1962, **76**, Whole No. 551.

Kohlberg, L. Early education: A cognitive-developmental view. *Child Developm.*, 1968, **39**, 1013-1062.

Minor, J. B. *Intelligence in the United States*. New York: Springer, 1957.

Reed, T. E. Caucasian genes in American Negroes. Unpublished manuscript, March, 1969.

Scott, J. P., & Fuller, J. L. *Genetics and the social behavior of the dog*. Chicago: Univer. of Chicago Press, 1965.

Shields, J. *Monozygotic twins brought up apart and brought up together*. London: Oxford Univer. Press, 1962.

Skeels, H. M. Adult status of children with contrasting early life experiences: A follow-up study. *Child Developm. Monogr.*, 1966, **31**, No. 3, Serial No. 105.

Skeels, H. M., & Dye, H. B. A study of the effects of differential stimulation on mentally retarded children. *Proc. Addr. Amer. Ass. Ment. Defic.*, *1939*, **44**, 114-136.

Tanner, J. M. Earlier maturation in man. *Sci. Amer.*, 1968, **218**, 21-28.

Tanner, J. M. The trend towards earlier physical maturation. In J. E. Meade & A. S. Parkes (Eds.), *Biological aspects of social problems*. New York: Plenum Press, 1965. Pp. 40-66.

Tuddenham, R. D. Soldier intelligence in World Wars I and II. *Amer. Psychol.*, 1948, **3**, 54-56.
Tuddenham, R. D. Psychometricizing Piaget's methode clinique. Paper read at Amer. Educ. Res. Ass., Chicago, February, 1968.
Vernon, P. E. Environmental handicaps and intellectual development: Part II and Part III. *Brit. J. educ. Psychol.*, 1965, **35**, 1-22.

Notes on Contributors

CARL BEREITER is Professor in the Departments of Applied Psychology and of Curriculum at the Ontario Institute for Studies in Education. Formerly Professor of Special Education in the Institute for Research on Exceptional Children at the University of Illinois, Professor Bereiter is currently concerned with early childhood education. He is co-author, with S. Englemann, of *Teaching Disadvantaged Children in the Preschool* (1966).

WILLIAM F. BRAZZIEL is Director of General Education at Norfolk State College. He is a psychologist and taught previously at Southern University in Baton Rouge, Louisiana, and at Central State College, in Wilberforce, Ohio. In addition to writing many articles, he has done research on training the hard-core unemployed.

LEE J. CRONBACH is Vida Jacks Professor of Education at Stanford University and was previously a professor at the University of Illinois. Professor Cronbach is primarily interested in educational psychology. His latest book, now in press, is *The Essentials of Psychological Testing* (3rd edition).

JAMES F. CROW is Professor of Genetics at the University of Wisconsin in Madison, where he was previously Acting Dean of the School of Medicine. Currently concerned with population genetics, he is the author of *Genetics Notes* (1966).

DAVID ELKIND is Professor of Psychology and Director of Graduate Training in Developmental Psychology at the University of Rochester. He is primarily concerned with cognitive development in children and adolescents, with parent-child relations, and with delinquency. Professor Elkind is the editor of *Six psychological studies by Jean Piaget* (1967) and co-editor, with J. H. Flavell, of *Studies in cognitive development: Essays in honor of Jean Piaget* (1969).

J. McV. Hunt is Professor of Psychology and Education at the University of Illinois and the former director of the Coordination Center of the National Laboratory of Early Childhood Education. His major interest lies in the experiential roots of intelligence and motivation. Professor Hunt is the author of *Intelligence and Experience* (1961) and of *The Challenge of Incompetence and Poverty,* to be published in either the summer or fall of 1969.

Arthur R. Jensen, Professor of Educational Psychology at the University of California at Berkeley, is the author of numerous articles on individual differences in human learning, serial learning, and cultural, developmental, and genetic determinants of intelligence and learning ability. He is a co-editor, with M. Deutsch and I. Katz, of *Social Class, Race, and Psychological Development* (1968).

Jerome Kagan is Professor of Developmental Psychology at Harvard University. Formerly on the faculties at Ohio State University and Antioch College, he is interested in early childhood development. Among his publications are, with G. S. Lesser, *Contemporary Issues in Thematic Apperception Methods* (1961), with H. A. Moss, *Birth to Maturity* (1962) and, with P. H. Mussen and J. J. Conger, *Child Development and Personality* (3rd ed., 1969).

Harvard Educational Review

VOLUME 39 1969 NUMBER 3

ARTICLES

*Ed. Note: Previously printed on p. 209 of *Harvard Educational Review*, Reprint Series No. 2, p. 673 this committee print.

Social Allocation Models of Intelligence: A Methodological Inquiry

RICHARD J. LIGHT and PAUL V. SMITH,
Harvard University

The authors consider Professor Jensen's hypothesis that inherited factors may be "implicated" in observed racial differences in measured intelligence. They argue that even if one chooses to accept Professor Jensen's estimates of the proportion of variance in intelligence accounted for by heredity, environment, and their interaction, his hypothesis is not substantiated by his own data. They go on to say that the parameter estimates are highly suspect, given the small sample size of the twin studies and the way disparate studies were combined. The authors simulated, on a computer, the process of studying twins and found that the statistical procedures employed in these studies of intelligence yield quite unstable estimates. In particular, the estimate of the interaction effect is quite unreliable, both because of sample size, and because Jensen chose a statistical model which would attribute some interaction to the main variables—heredity and environment. Finally, the authors propose that the studies of intelligence reported by Professor Jensen ignore the reality of feedback loops, initiated by physical differences, and enhanced by processes of social differentiation in our society.

Reaction to the recent article by Arthur R. Jensen[1] has been both harsh and predictable. However, much of it has been political, and therefore convincing or unconvincing, depending upon one's politics. Jensen's position is that it is "... a

[1] Arthur R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?," *Harvard Educational Review*, XXXIX (Winter, 1969), 1-123.

Harvard Educational Review Vol. 39 No. 3 Summer 1969

Models of Intelligence

RICHARD J. LIGHT AND PAUL V. SMITH

not unreasonable hypothesis that genetic factors are strongly implicated in the average Negro-white intelligence difference" (p. 82). We believe his article contains serious technical errors, both of commission and omission, and we therefore present several statistical arguments which substantially weaken the plausibility of Jensen's hypothesis. We argue that even a tentative acceptance of this hypothesis is not justified by a careful analysis of available data.

Three strategies are available for developing a case against Jensen's postulated relation between race and inherited intelligence.

1. Accept Professor Jensen's model, and his estimates of the model parameters, but show that non-genetic disparities would still account for most or all of the 15 point IQ difference between the races. This explanation does not implicate genetic racial differences and yet is consistent with his estimates of the model parameters.
2. Accept Professor Jensen's mathematical model as a satisfactory descriptive construct, but question his parameter estimates as being either unreliable or incorrect.
3. Reject Professor Jensen's model as inappropriate because of its failure to incorporate relationships which clearly exist among variables in the real world as we know it.

In our analysis, we will adopt each of these three strategies, and demonstrate that they all lead to inferences in direct contradiction to Jensen's suggestion that genetic differences are implicated in observed black-white IQ differences.

Four Major Results of Our Analysis:

First, we accept Professor Jensen's estimate of the proportion of IQ variability explained by hereditary versus environmental factors. Using his proportions, with no interaction effects, we show that more than half of the 15 point difference in mean IQ between races is explained by the differential allocation of the races to social conditions. Second, we demonstrate that if the interaction between genetic and environmental effects accounts for only one percent of the total IQ variance, then a 15 point difference between means occurs even if the racial genetic distributions were identical. Since Jensen's data suggest that the proportion of interaction variance is one percent, this analysis is entirely consistent with his data. Third, we present the results of computer simulations which establish the instability of Jensen's estimation procedures, thus throwing doubt on

the reliability of his parameter estimates. We conclude that if the usual statistical criteria for significance are used, Jensen's data are not inconsistent with the presence of a moderate amount of genetic-environmental interaction. Fourth, we show that if moderate amounts of interaction exist, mean IQs of blacks and whites could easily differ by more than 15 points, despite identical genetic distributions in the two races. These are our statistical arguments, all of which assume that Jensen's model and his estimation procedures are conceptually tenable. Later in the paper, however, we question the quality and the consistency of Jensen's data. Finally, we conclude by suggesting that current genetic-environment models of intelligence ignore important social processes and must be revised.

Jensen's Estimates:

Jensen uses a components of variance model to estimate the proportion of total IQ variability attributable to genetic factors, environmental factors, and their interaction.[2] It is crucial to note, however, that Jensen estimated the three components in a very particular way.[3] From studies of monozygotic twins (MZ), he derives an estimate of the genetic proportion as .75. Then, from studies of un-

[2] The original components of variance model of the type Jensen uses was proposed by C. Eisenhart, "The Assumptions Underlying Analysis of Variance," *Biometrics*, 3, 1947, 1-21. This model is now called one of two names by statisticians; a Model II random effects model, or a Graybill Model 5 Components of Variance. An excellent and very readable treatment of this model appears in W. L. Hays, *Statistics for Psychologists* (New York: Holt, Rinehart & Winston). A detailed treatment of the statistical assumptions underlying this model is available also in S. L. Crump, "The Present Status of Variance Components Analysis," *Biometrics*, 7, 1951, 1-16. The intraclass correlation coefficients used to estimate the various proportions of variance differ from the well known product moment correlation coefficients, although Jensen never makes this distinction directly. The intraclass correlation was developed by Fisher, and clearly described in his classic, R. A. Fisher, *The Design of Experiments* (London: Oliver-Boyd, 1935).

[3] Jensen, in his article, reports intraclass correlations taken from MZ twin and foster pair studies. He nowhere indicates that these correlations depend upon and are derived from the special assumptions of a random effects analysis of variance (ANOVA). This Model II ANOVA requires that the main effects have been sampled at random from the underlying dimension, which is the population of genotypes for the MZ studies, and the relevant array of possible environments for the foster pair studies. Further it requires that the 'other' dimension in each case has been allocated at random within, as well as between, pairs. We actually would want to estimate a two-way ANOVA, with both genetics and environment separately replicated but crossed randomly—but no such design is possible. Instead, two one-way ANOVAs are used to estimate the main effects, and the interaction variance that would have been found directly in a two-way study is estimated by subtraction. Jensen essentially uses three parameters as additive components for this model: V_G (genetic variance), V_E (environmental variance), and V_I (interaction variance). The estimates are obtained from the pair studies, and thus based upon single replication within groups. These components sum to the total phenotypic variance in the population, V_P. Further, for the MZ twin studies, the Model II ANOVA produces a direct link between the intraclass correlation coefficient *r*, the heritability coefficient H, and the ratio of V_G to V_P. The link is that $H = r = V_G/V_P$.

related foster children raised together, he estimates the environment proportion as .24. This leaves .01 for the third component, the genetic-environment interaction.

We should point out that an estimate of a percent of variance does *not* translate directly into any explanation of a difference between racial mean IQs. Otherwise, the analysis would end right here, and we could conclude that 75 percent of the 15 point gap is due to genetic factors. Such reasoning is simply incorrect.

Social Allocation Models:

We call our explanation of the racial differences in mean IQ a social allocation model. By social allocation we mean a process whereby members of different racial groups are assigned to environments non-randomly. This model differs from the classic environmentalist position because it grants that individual differences in IQ (although not racial differences) are largely genetic. With respect to the racial differences, the 25 percent non-genetic component accounts for all the observed difference. We differ from the interactionists in believing that the majority of the variation in intelligence can be separated into additive genetic and environmental components. We have found that the presence of a tiny interaction component can explain large differences between group means.

Other critics of Professor Jensen's analysis have already pointed out that given the structure of today's society, blacks and whites are exposed, or 'socially allocated,' to essentially different environments. For example, in his chapter in the excellent book edited by Deutsch, Katz, and Jensen, I. I. Gottesman points out that "In the light of what has been said in the introduction to this book and the literature on the effects of stimulation in early infancy (Caster, 1961; Riesen, 1961; Thompson and Schaefer, 1961), there is every reason to doubt that a typical Negro infant is reared in a typical white infant's natural habitat."[4]

At several points in his analysis, however (pp. 74-78, 83, 87), Jensen seems to accept one or another of the current indexes of socio-economic status as being a comparable measure of the social condition of blacks and whites. Nearly all such indexes are weightings of education (in number of years completed), income, and occupation (on one or another rating of prestige). These are obviously important dimensions within the social life of our nation, and the two races differ

[4] I. I. Gottesman, "Biogenetics of Race and Class," in *Social Class, Race, and Psychological Development*, ed. by M. Deutsch, I. Katz, and A. R. Jensen (New York: Holt, Rinehart & Winston, 1968).

considerably in their placement on these three dimensions.[5] But the crucial point is that our SES measures do not work identically for both races. Black and white families with identical incomes do not have the same economic options open to them—housing discrimination is only the most obvious reason for assuming that a dollar is worth less to a black family than to a white one. Further, equal numbers of years of education do not imply equal educations; nor do they imply equal access to further benefits, such as jobs or income levels.

That black engineers, college presidents, physicians, and teachers all receive less than similarly employed whites is well known; however, even black postal clerks who are high school graduates earn on the average $250 less per year than their white peers.[6] Other racially related inequities, such as in promotion possibilities, nepotistic privileges, community prestige, are all also lost within our conventional indexes. Equal standings on our measures of social status do not imply equal statuses—at least not for men of different races.

We thus cannot draw conclusions from differences in IQ among blacks and whites "of equal SES" simply because equal SES scores still imply more restricted life-chances for blacks. Jensen, in his Table 3 (page 83), gives data on the prevalence of children with IQs below 75 for each race, holding measured SES constant. He is incorrect in assuming that the actual environments are identic l.[7]

On the other hand, an SES index does provide a rough suggestion of the existing range of variation in social conditions. Jensen used SES groupings as a measure of *equality* between whites and blacks—we will use them as a measure of the *disparity* between the races. Since SES measures understate that disparity, we are making an inherently conservative estimate of environmental effects.

We promised to use both Jensen's model and his parameter estimates to illustrate our social allocation model; we now present this investigation and show that non-genetic differences can explain the observed differences in mean IQ.

[5] Whether they are also the relevant dimensions upon which to measure environments with regard to their impacts upon IQ is uncertain.

[6] U. S. Census, 1960, *Subject Reports, Occupation by Earnings and Education*, Report PC(2)-7B, U. S. G.P.O., Washington, D. C., pp. 102, 103.

[7] The SES scale does not even behave similarly for the two races in its gross properties—always a strong signal that we are not measuring the same thing in each population. The U. S. Census SES scale was built from three components (education, income, and occupation, as usual); persons were registered as being "consistent" in SES if they had scores on each of the three underlying 100 point scales which all fell within 20 points of each other. The median overall SES for all whites was 55.2, while the median SES for consistent whites was 11.8 points *higher* at 67.0. For blacks, the median SES was 28.5, but the consistent blacks had scores 15.9 points *lower*, at 12.6. Note that this cannot be a floor-and-ceiling effect, since the displacement is in the wrong direction in both cases.

Models of Intelligence
RICHARD J. LIGHT AND PAUL V. SMITH

Exploring Racial Differences via Two-way Tables:

We will use the well-known format of the two-way layout to examine socially allocated racial differences in IQ.

Let us design a 12 by 10 table: twelve rows and ten columns. Suppose we let the rows represent some measure of environment, which is divided into twelve types, and the columns represent some measure of genetic endowment, which will be divided into ten types. We then have a table where each of the 120 cells represents a particular genetic-environment combination. Further, we may assign two numbers to each cell; one number corresponding to the mean IQ of all persons falling into that cell, and the second number giving the proportion of people from some population falling into the cell. If we have all these data available to us, then we may compute the mean IQ for any population by simply computing an overall weighted mean from the 120 cell means, where each cell mean is weighted by the proportion of people in the population falling into that cell.

If we are going to use this approach to compare black-white IQ differences, then the first order of business is to develop separate tables for each race. However, we will argue that blacks and whites have no differences in genetic endowments, which requires that the tables for each race have an identical genetic distribution over the ten columns. Therefore, we may begin by assigning a column mean to each of the ten columns.

Our choice of the number of columns to use was fairly arbitrary, but our assignment of column means is not at all arbitrary. We will begin by accepting Jensen's estimate that the genetic component of variance is 75 percent of the total phenotypic IQ variance. This then requires, if we use an IQ total variance of 225 as a reasonable estimate for both populations combined, that the variance over the ten genetic categories be 75 percent of 225, or 168.75. Accepting also Jensen's contention that genetic endowments are approximately normally distributed, we may then divide the population of each race into tenths, so that each of the ten genetic categories contains one tenth of the population. To determine the column means, finally, we simply find the mean value of each tenth of a normal distribution which has an overall mean of 100.00 and a variance of 168.75. These means are given in Table A, which gives not only the column means when the genetic proportion of phenotypic variance is .75, but also the means when this proportion is .70 and .80, in case the reader wishes to try alternative partitions of variance.

The determination of the twelve row means is more complicated, because although we posited that whites and blacks are identically distributed over the

TABLE A

Percentage Distributions by Race, and Mean IQ Scores, for the Genetic Categories. Means are given for the Cases when the Proportions of the Total IQ Variance (225) accounted for by Genetic Effects are .70, .75, and .80.

Genetic Category:		*Percent of Population in Each Category*[a] *Whites:*	*Blacks:*	*Mean of Each Category if the Proportion of Variance is:* *.70*	*.75*	*.80*
1.	(low)	10.0	10.0	78.01	77.02	76.49
2.		10.0	10.0	86.88	86.30	86.04
3.		10.0	10.0	91.52	91.09	90.88
4.		10.0	10.0	95.11	94.85	94.82
5.		10.0	10.0	98.41	98.27	98.20
6.		10.0	10.0	101.59	101.73	101.80
7.		10.0	10.0	104.89	105.15	105.18
8.		10.0	10.0	108.48	108.91	109.12
9.		10.0	10.0	113.12	113.70	113.96
10.	(high)	10.0	10.0	121.99	122.98	123.51

[a] The two races are assumed to have identical genetic distributions, with one tenth of each race falling into each of the ten genetic categories. A larger number of genetic categories (the choice of ten was arbitrary) *increases* the effects reported in this paper, but only slightly.

ten columns (genetic effects), the two races are clearly distributed differently over the twelve rows, representing environment effects.

First, we must decide how to measure relative environments for blacks and whites. Assume that the SES of the families of white and black children under 14 is a rough indicator of the environments in which the children were raised. Since the estimate of the environment's contribution to IQ was derived from studies of white children, the white proportions in each of the 12 SES groupings provided by the U.S. Census will be used to estimate the mean values for these categories.[8] Then we will ask what the effect of the environmental dimension would be upon a group allocated as the black children are. Because SES understates the disparity between blacks and whites, this procedure *underestimates* the impact of environmental deprivation upon black children.

We will use an estimate of the proportion of phenotypic variance in IQ due to environment of 25 percent. (Thus, we are allowing exactly zero interaction between genetic effects and environmental effects.) 25 percent of a total IQ variance

[8] U. S. Census, 1960, *Subject Reports, Socioeconomic Status*, Report PC(2)—56, U. S. G. P. O., Washington, D. C. p. 50.

of 225 yields an environment component variance of 56.25, which means that our overall environment axis will be represented by a normal distribution with overall mean 100.00 and variance 56.25.

Some readers may find it less appealing to assume a normal distribution for environments than for genotypes. Our perception of how the real world works suggests that there are very disabling environments to which far more children are subjected than would occur within a normal distribution—even among the white population, but especially among the blacks. It can also be argued that prejudice afflicts most black children in a way no white child experiences. If either of the above intuitions is correct, our assumption understates the deprivation of the black population, and thus the case against Professor Jensen's position is stronger than we present. Readers who are unpersuaded by our model for environments should be equally unpersuaded by Professor Jensen's proportions of variance and his conclusions about racial differences.

Since blacks and whites are not identically distributed over the twelve environment categories, we will not break this new normal distribution into equal

TABLE B

Percentage Distributions by Race, and Mean IQ Scores, for Environmental (SES) Categories. Means are given for the Cases when the Proportions of the Total IQ Variance (225) accounted for by Environment are .15, .20, and .25.

Environment Category:		*Percent of Population in Each Category*[*]		*Mean of Each Category if the Proportion of Variance is:*		
		Whites:	*Blacks:*	*.15*	*.20*	*.25*
1.	(low)	2.0	15.8	86.11	83.80	81.96
2.		4.4	20.2	89.89	88.29	86.96
3.		3.2	10.0	91.86	90.61	89.46
4.		4.1	10.1	93.09	92.02	91.06
5.		10.9	16.9	94.93	94.09	93.42
6.		14.3	12.0	97.23	96.79	96.39
7.		16.3	7.3	99.56	99.49	99.42
8.		15.6	4.0	101.94	102.25	102.51
9.		7.1	1.3	103.81	104.40	104.62
10.		5.9	.9	105.07	105.86	106.51
11.		9.3	1.1	107.02	108.16	109.02
12.	(high)	6.8	.5	111.26	112.93	114.30

[*] From p. 50, U. S. Census of Population, 1960, *Subject Reports: Socioeconomic Status*. Final Report PC(2)—5C. U. S. Government Printing Office, Washington, D. C.

twelfths. Rather, we will use the proportion of whites in each of the twelve categories to determine what proportion of the environment distribution should be allocated to each category, and we thus find the mean for each of the twelve rows. These means, as well as the proportion of whites in each category, are presented in Table B, which also gives similarly derived means for an environmental variance component of 15 percent and 20 percent, to be used a bit later.

We thus have found twelve row means and ten column means; the additive ANOVA model, with no interaction effects present, then enables us to find a mean value for each of the 120 cells in our original table. These cell means are given in Table 1A.

TABLE 1A

Mean IQ Score for each Combination of the Ten Genetic and Twelve Environmental Categories, when 75 percent of the Total Variance is Genetic, 25 percent is Environmental, and There are no Interaction Effects Present.

Environment (SES) Category (Rows)	*Genetic Category (Columns)*										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Row Mean*
1	58.98	68.26	73.05	76.81	80.23	83.52	86.90	90.69	95.42	104.71	81.96
2	63.98	73.26	78.05	81.81	85.23	88.52	91.90	95.69	100.42	109.71	86.96
3	66.48	75.76	80.55	84.31	87.73	91.02	94.40	98.19	102.92	112.21	89.46
4	68.08	77.36	82.15	85.91	89.33	92.62	96.00	99.79	104.52	113.81	91.06
5	70.44	79.72	84.51	88.27	91.69	94.98	98.36	102.15	106.88	116.17	93.42
6	73.41	82.69	87.48	91.24	94.66	97.95	101.33	105.12	109.85	119.14	96.39
7	76.44	85.72	90.51	94.27	97.69	100.98	104.36	108.15	112.88	122.17	99.42
8	79.53	88.81	93.60	97.36	100.78	104.07	107.45	111.24	115.97	125.26	102.51
9	81.64	90.42	95.71	99.47	102.89	106.18	109.56	113.35	118.08	127.37	104.62
10	83.53	92.81	97.60	101.36	104.78	108.07	111.45	115.24	119.97	129.26	106.51
11	86.04	95.32	100.11	103.87	107.29	110.58	113.96	117.75	122.48	131.77	109.02
12	91.32	100.60	105.39	109.15	112.57	115.86	119.24	123.03	127.76	137.05	114.30
Column Mean	77.02	86.30	91.09	94.85	98.27	101.73	105.15	108.91	113.70	122.98	

When the cell mean IQ scores in this table are multiplied by the proportions taken from the corresponding cell in Table 1B, and the products are summed, the result is the average IQ for the white population. That average IQ equals 100. When the same procedure is followed using the proportions for the black population, taken from Table 1C, the result is an average IQ for black children of 91.26—in spite of the fact that both races have identical genetic distribution.

TABLE 1B

Percent of the White Population Allocated to Each Combination of the Ten Genetic and Twelve Environmental Categories.

Environment (SES) Category (Rows)	*Genetic Category (Columns)* 1	2	3	4	5	6	7	8	9	10	Row %
1	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	2.0
2	.44	.44	.44	.44	.44	.44	.44	.44	.44	.44	4.4
3	.32	.32	.32	.32	.32	.32	.32	.32	.32	.32	3.2
4	.41	.41	.41	.41	.41	.41	.41	.41	.41	.41	4.1
5	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	1.09	10.9
6	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	1.43	14.3
7	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	1.63	16.3
8	1.56	1.56	1.56	1.56	1.56	1.56	1.56	1.56	1.56	1.56	15.6
9	.71	.71	.71	.71	.71	.71	.71	.71	.71	.71	7.1
10	.59	.59	.59	.59	.59	.59	.59	.59	.59	.59	5.9
11	.93	.93	.93	.93	.93	.93	.93	.93	.93	.93	9.3
12	.68	.68	.68	.68	.68	.68	.68	.68	.68	.68	6.8
Column %	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	100.0

The within-cell entries which appear in the table are those assumed by the twin studies. In fact, as long as the row and column constraints are met, any other set of cell entries would result in the same overall mean IQ.

Now, finally, we are ready to compare the overall black and white mean IQs. Table 1B shows how whites are allocated to each of the 120 cells. Notice that this allocation table has ten percent of all whites falling into each of the ten genetic categories. Combining the table of cell means with the white allocations, we find that white children under 14 have an average IQ of 100.00, as Jensen reports.[a]

Turning to the table of black allocations (Table 1C), we find that the row totals are different from those for whites. These black row totals correspond to the proportion of blacks falling into the same twelve environment categories as defined by the U.S. Census Bureau. As we all could have guessed, the blacks cluster more towards the low socio-economic categories than do the whites. But the blacks are also assigned with no interaction or covariance, and ten percent of all black children under 14 are assigned to each of the ten genetic categories.

[a] Of course, the overall mean IQ could have been computed from only the row marginals, with each row mean weighted by the proportion of whites assigned to that row. But this is only the case when the table of cell means has no interaction effects present.

TABLE 1C

Percent of the Black Population Allocated to Each Combination of the Ten Genetic and Twelve Environmental Categories.

Environment (SES) Category (Rows)	*Genetic Category (Columns)*										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Row %*
1	1.58	1.58	1.58	1.58	1.58	1.58	1.58	1.58	1.58	1.58	15.8
2	2.02	2.02	2.02	2.02	2.02	2.02	2.02	2.02	2.02	2.02	20.2
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	10.0
4	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	10.1
5	1.69	1.69	1.69	1.69	1.69	1.69	1.69	1.69	1.69	1.69	16.9
6	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	12.0
7	.73	.73	.73	.73	.73	.73	.73	.73	.73	.73	7.3
8	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40	4.0
9	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13	1.3
10	.09	.09	.09	.09	.09	.09	.09	.09	.09	.09	.9
11	.11	.11	.11	.11	.11	.11	.11	.11	.11	.11	1.1
12	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.5
Column %	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	100.0

The within-cell entries which appear in the table are those assumed by the twin studies. In fact, as long as the row and column constraints are met, any other set of cell entries would result in the same overall mean IQ.

Combining the cell mean table with the black allocation table, the overall black mean IQ computes to 91.26. Thus, we see that if Jensen's essential breakdown of .75 and .25 for the G and E proportions of effects on IQ is used, and if U.S. Census data on the differential social allocation of the two races to socio-economic categories are used, *more than half of the mean difference in IQs between races is explained, although the races have identical genetic distributions.*

Let us pause for a moment and see what lowered the mean IQ of the black children. Within each racial group separately, most of the variation among children is fundamentally genetic: this follows from the way in which the marginal means were computed, but can be most easily observed by noticing that more than half of both groups falls into 4 of the 12 rows in their tables. In other words, most of the children have environments quite similar to others of their race. But the black children all cluster toward the lower SES levels while the white children cluster

toward the middle. Even though three quarters of the variance within each race is genetic, the variance between the races is wholly environmental.

We therefore suggest that whites have a distribution over environment types which has a relatively small variance, and blacks similarly have an environmental distribution with small variance. However, the variance *between* the white and black allocations in large. Thus, with only non-genetic differences of 25 percent to separate the races, an expected difference in mean IQs of nearly 9 points is found.

Introducing Interaction:

Our analysis thus far has taken the most favorable case for Professor Jensen's point of view, by using the proportion of .75 for genetic effects, and attributing

TABLE 2A

Interaction Effects on Each Combination of the Ten Genetic and Twelve Environmental Categories, when 1 percent of the Total Variance in IQ is due to Interaction.

Environment (SES) Category (Rows)	*Genetic Category (Columns)* 1	2	3	4	5	6	7	8	9	10	*Row Sum*
1	—7.34	—4.24	1.97	1.51	1.30	1.45	2.18	1.45	.99	.72	0.0
2	2.26	—1.64	—9.43	—2.89	1.90	2.05	2.78	2.05	1.59	1.32	0.0
3	.59	.61	.65	—1.64	—2.01	.22	.70	.22	.26	.40	0.0
4	.59	.61	.65	.36	—2.01	—1.78	.70	.22	.26	.40	0.0
5	2.36	2.46	2.67	2.21	2.00	—1.85	—11.12	—1.85	1.69	1.42	0.0
6	.59	.61	.65	.36	—.01	.22	.70	—1.78	—1.74	.40	0.0
7	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
8	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
9	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
10	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
11	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
12	.16	.26	.47	.01	—.20	—.05	.68	—.05	—.51	—.78	0.0
Column Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

The interaction values shown in this table are arbitrary, in that there are an infinite number of such tables, each displaying about 1 percent of the total variance in IQ.

This table was selected to illustrate the consequences of relatively small amounts of interaction, and has no direct empirical basis. Rows and columns may not add to zero because of rounding. Table 3A was similarly constructed, but has an interaction proportion of 10 percent.

the entire remaining .25 to environment effects. Jensen's actual breakdown of proportions was .75 and .24 for genetic and environment effects respectively, with the .01 remainder allowed for interaction. We now examine the possible influence of this tiny interaction component on the overall racial mean IQs.

A .01 interaction effect permits us to introduce an interaction variance of 2.25 into our two-way layout, while dropping the environment variance to 24 percent of 225, or 54.0. Thus, we may modify the cell mean IQs in each of the 120 cells, so long as we satisfy our variance constraint. Our modifications for this case appear in Table 2A. This table of interaction effects gives the particular effects we chose to inject; there are an infinite number of possible interaction effect tables which would satisfy the variance constraints. We note, however, that our particular table satisfies Jensen's parameter estimates, which was our primary concern.

Table 2B gives the adjusted cell mean IQ values, after the interaction effects were included. Finally, Table 2C of black allocations illustrates how our social

TABLE 2B

Mean IQ Score for Each Combination of the Ten Genetic and Twelve Environmental Categories, when 75 percent of the Total Variance is Genetic, 24 percent is Environmental, and 1 percent is due to the Presence of Interaction Effects.

Environment (SES) Category (Rows)	*Genetic Category (Columns)*									
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1	51.64	64.02	75.02	78.32	81.53	84.97	89.08	92.14	96.41	105.43
2	66.24	71.62	68.62	78.92	87.13	90.57	94.68	97.74	102.01	111.03
3	67.07	76.37	81.20	82.67	85.72	91.24	95.10	98.41	103.18	112.61
4	68.07	77.97	82.80	86.27	87.32	90.84	96.70	100.01	104.78	114.21
5	72.80	82.18	87.18	90.48	93.69	93.13	87.24	100.30	108.57	117.59
6	74.00	83.30	88.13	91.60	94.65	98.17	102.03	103.34	108.11	119.54
7	76.60	85.98	90.98	94.28	97.49	100.93	105.04	108.10	112.37	121.39
8	79.69	89.07	94.07	97.37	100.58	104.02	108.13	111.19	115.46	124.48
9	81.80	91.18	96.18	99.48	102.69	106.13	110.24	113.30	117.57	126.59
10	83.69	93.07	98.07	101.37	104.58	108.02	112.13	115.19	119.46	128.48
11	86.20	95.58	100.58	103.88	107.09	110.53	114.64	117.70	121.97	130.99
12	91.48	100.86	105.86	109.16	112.37	115.81	119.92	122.98	127.25	136.27

When the cell means in this table are weighted by the proportions shown for the black population (in Table 2C) the average IQ is reduced to 86.81. Weighted by the white distribution from Table 1B, the result remains 100.0. Since the whites and blacks are shown as genetically identical, the disparity is due only to the generally disadvantageous and specifically malicious allocation of environments to the black children.

TABLE 2C

Malicious Allocation of Blacks to Genetic-Environmental Combinations which have Negative Interaction Effects Present. The Black Population has the Same Genetic Distribution as the White Population.

Environment (SES) Category (Rows)	*Genetic Category (Columns)*										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Row %*
1	8.60	4.72	.18	.18	.18	.18	.18	.18	.18	1.12	15.8
2	.18	4.00	8.57	5.64	.18	.18	.18	.18	.18	.88	20.2
3	.17	.17	.17	3.13	5.16	.17	.17	.17	.17	.54	10.0
4	.17	.17	.17	.17	3.50	5.15	.17	.17	.17	.27	10.1
5	.18	.18	.18	.18	.18	3.66	8.61	2.86	.18	.68	16.9
6	.17	.17	.17	.17	.17	.17	.17	5.87	4.45	.47	12.0
7	.16	.16	.16	.16	.16	.16	.16	.16	4.08	1.96	7.3
8	.13	.13	.13	.13	.13	.13	.13	.13	.13	2.80	4.0
9	.08	.08	.08	.08	.08	.08	.08	.08	.08	.59	1.3
10	.06	.06	.06	.06	.06	.06	.06	.06	.06	.34	.9
11	.07	.07	.07	.07	.07	.07	.07	.07	.07	.46	1.1
12	.04	.04	.04	.04	.04	.04	.04	.04	.04	.14	.5
Column %	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	10.0	100.0

allocation model works. The proportion of blacks in each of the twelve rows is unchanged from before; however, within rows, the blacks are more heavily weighted in those columns which also have negative interaction effects. Thus, examining the black allocation table, we find that our social allocation of blacks has been 'malicious,' in that the standing of any black in his genetic category is likely to be higher than his standing on environment.

If we weight the cell IQ means in Table 2B by the black allocations to genetic-environment combinations, we find that although the white mean IQ is unchanged at 100.00, the black mean has now dropped to 86.81. Once again, we recall that we have both kept the genetic distributions of blacks and whites identical and used the exact parameter estimates obtained by Jensen. Thus, we see that the observed difference in IQ between the races is essentially entirely accounted for by our social allocation approach, with no genetic differences being present.

As a final analysis of this type, we investigated what would happen to the mean black IQ if we kept the genetic component at .75, while reallocating the remainder so that the environment variance component was .15, and the interaction

proportion .10. The results of this analysis appear in Table 3. Keeping the allocation of blacks to cells identical to the allocation of Table 2, the black mean drops to 82.59. We may therefore conclude that with an interaction component of variance somewhere between .01 and .10, the black mean IQ may be expected to be approximately 85, even though blacks are distributed identically with whites over the genetic categories.

Summing Up:

Our models show how a large difference between black and white mean IQs may be explained not by the hypothesis of genetic differences between races, but rather by the non-genetic differences in allocation of blacks and whites to different environments, and environment-genetic combinations. In terms of our 12 by 10 tables, Jensen has argued that the difference in mean IQs may be attributable to a differential assignment of whites and blacks to the *columns* of the table, with blacks clustered relatively more to the leftward columns than are whites. We believe, rather, that the differential allocation occurs on the *rows*, with blacks more frequently assigned to the upper rows than are whites. By combining our argument with the malicious allocation effects on blacks of interaction, the IQ differences are readily explainable without any resort whatever to the hypothesis of genetic differences between races.

The Parameter Estimates:

So far we have assumed that Professor Jensen's model and his parameter estimates are approximately correct. We now turn our attention to examining the validity of Jensen's particular estimates of P_G (the proportion of variance in measured IQ due to genetic factors), P_E (the proportion of variance due to environment), and P_I (the proportion of variance from interaction).

It is important to question the exact values of these parameters because in a society which allocates whites and blacks with similar genetic endowments differentially to environments, a small amount of interaction, P_I, can go a long way toward allocating people with certain genetic endowments to a favorable environment or an unfavorable environment for those particular endowments. And if a society chooses to take advantage of the genetic-environment interaction by allocating persons of one race in a positive way, while allocating persons of another race in a negative way, then that society can easily build in differences in measured intelligence where there were none at birth. Note we are not arguing

that no individual differences in intelligence exist. We agree with Jensen that these differences are present. Where we disagree with Jensen is with his specific argument suggesting the presence of *racial* differences in genetic endowments. It is here that our interaction argument becomes important, and because of its importance, we digress for a moment to consider its implications.

Implications of Interaction:

We pointed out at the beginning of this article that Jensen's components of variance approach to estimating intelligence uses a Model II ANOVA. Statistical interaction, as imbedded in the Model II ANOVA, relates to a gain or loss in intelligence due to a particular combination of E and G placements. This gain or loss is not predictable from both E and G considered separately in an additive sense. The Model II ANOVA, however, assumes that the cell values, weighted by the proportion of the population in each cell, sum to zero across both variables. Actually, the Model II not only assumes the weighted terms sum to zero; *it insures that they do by attributing any residual to the main effects of environment and*

TABLE 3A

Interaction Effects on Each Combination of the Ten Genetic and Twelve Environmental Categories, when 10 percent of the Total Variance in IQ is due to Interaction.

Environment (SES) Category (Rows)	*Genetic Category (Columns)*										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	
1	—23.66	—11.03	4.34	4.34	4.34	4.34	4.34	4.34	4.34	4.34	0.0
2	4.46	—3.96	—19.90	—1.72	4.10	3.91	4.46	3.91	2.82	1.92	0.0
3	3.16	2.74	2.80	—8.02	—11.20	2.61	3.16	2.61	1.52	.62	0.0
4	3.16	2.74	2.80	1.98	—7.20	—11.39	3.16	2.61	1.52	.62	0.0
5	5.16	4.74	4.80	3.98	4.80	—3.39	—22.84	—3.39	3.52	2.62	0.0
6	3.16	2.74	2.80	1.98	2.80	2.61	3.16	—11.39	—8.48	.62	0.0
7	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
8	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
9	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
10	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
11	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
12	.76	.34	.40	—.42	.40	.21	.76	.21	—.88	—1.78	0.0
Column Sum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

TABLE 3B

Mean IQ Score for Each Combination of the Ten Genetic and Twelve Environmental Categories, when 75 percent of the Total Variance is Genetic, 15 percent is Environmental, and 10 percent is due to the Presence of Interaction Effects.

Environment (SES) Category (Rows)	*Genetic Category (Columns)* 1	2	3	4	5	6	7	8	9	10
1	39.46	61.37	81.53	85.29	88.71	92.00	95.38	99.17	103.90	113.18
2	71.37	72.23	61.08	83.02	92.26	95.36	99.29	102.53	106.17	114.36
3	72.03	80.89	85.74	78.68	78.92	96.02	99.95	103.19	106.83	115.22
4	73.27	82.13	86.98	89.92	84.16	83.26	101.19	104.43	108.07	116.46
5	77.10	85.96	90.81	93.75	97.99	93.09	77.02	100.26	111.90	120.29
6	77.40	86.26	91.11	94.05	98.29	101.39	105.32	94.56	102.20	120.59
7	77.33	86.19	91.04	93.96	96.22	101.32	105.25	108.49	112.13	120.52
8	79.71	88.57	93.42	96.36	100.60	103.70	107.63	110.87	114.51	122.90
9	81.58	90.44	95.29	98.23	102.47	105.57	109.50	112.74	116.38	124.77
10	82.85	91.21	96.56	99.50	103.74	106.84	110.77	114.01	117.65	126.04
11	84.79	93.65	98.50	101.44	105.68	108.78	112.71	115.95	119.59	127.98
12	89.04	97.90	102.75	105.69	109.93	113.03	116.96	120.20	123.84	132.23

Weighted by the proportional distribution of the black population given in Table 2C, this table results in a black average IQ of 82.59. The white population remains unaffected, with a mean of 100.0. Therefore a difference between the mean IQs of the two races of more than 15 points has occurred, yet the genetic potential of both races remains exactly equal.

genetic endowment. This means that, given a world in which a specific combination of genotype and environment always produced exactly the same measured IQ, we could, merely by changing the number of persons assigned to the specific genetic environment combinations, set the proportion of variance 'due to interaction' to various values. The statistical concept of interaction *enforced* by the Model II ANOVA, then always produces the *lowest possible* estimate of the percent of variance due to interaction. Indeed, in a world in which all the interaction terms were used, by selective assignment, to benefit one race, and to deprive another, *Model II ANOVAs carried out within the two groups separately would show no interaction variance in either case, but would translate the true interaction component into (in part) a displacement of the two groups on the genetic axis relative to one another.*

Thus, the importance of interaction is twofold: one, interaction components may in fact exist and if they do they must be correctly estimated or any findings

Models of Intelligence
RICHARD J. LIGHT AND PAUL V. SMITH

are worth little; two, the existence of interaction holds out the possibility of reallocating environments among persons so that all benefit and none are harmed. That is, interaction, in the theoretical sense, is an eminently 'political' variable. If no interaction exists, then that environment which is best for one child is also best for every other child. On the other hand, if interaction terms really appear in the effect table for genetics and environment, then some environments will be better for some children, and other environments will be best for other children. Given limits on the number of environments of one kind which we can produce, the presence of interaction allows us to benefit all children by reallocating the environments we do have.

Unreliability of P_I:

Based on the discussion above, we believe it is clear that the precision of the estimate of the interaction variance is crucial, if inferences concerning racial differences in genetic endowments are to be drawn. We now examine Jensen's estimation procedure.

The basic data from which Professor Jensen concluded that $P_I = .01$ appear in his Table 2 (p. 49). He notes that this value is obtained from four studies of monozygotic twins which yield a median r for IQ of .75, and five studies of unrelated children reared together, which give a median r for IQ of .24. From his components of variance model, the proportion of variance in IQ explained by the interaction component is $1.00 - .75 - .24 = .01$. The crucial point to note here is that the interaction proportion was not estimated directly from any data; rather, it was estimated from a linear combination of environment and genetic effects.

Now it is well known that the variance of an estimate which is a linear combination of two variables, is the sum of the variances for the two variables. (This assumes independently collected data for the estimates of P_E and P_G, which we will assume.) And since we have argued at some length that the value of the interaction component is critical, we decided to investigate the variance of Jensen's estimate quite closely.

In order to do this, we gave birth to approximately 540,000 twins, in a computer simulation of what Jensen's analysis would be like if done a great many times. Specifically, we wished to study the variance of the estimate of intraclass correlation coefficients of the type used by Jensen, when the coefficients are the median correlations taken from several studies. We thus generated sets of five twin studies, which came from a population with a specified P_E and P_G. Each of

the individual twin studies was based on somewhere between 30 and 60 twins, drawn at random from a uniform distribution. For each panel of five studies, the median r of the five r's was found. For each P_E and P_G combination, 200 such panels were generated. The variance of each set of 200 median r's was then computed, and the results are given in Table 4.

TABLE 4

Standard Errors of Estimates of Median Intraclass Correlation Coefficient r, from Simulating 200 Panels of 5 Studies, each Study Containing 30-60 Twins.

	r_G	r_E	r_I
coefficient	.80	.20	0
standard error	.0310	.0760	.0818
coefficient	.80	.10	.10
standard error	.0310	.0822	.0895
coefficient	.70	.20	.10
standard error	.0490	.0760	.0915
coefficient	.70	.10	.20
standard error	.0490	.0822	.0942
coefficient	.60	.20	.20
standard error	.0507	.0769	.0921

The remarkable property which these median estimates of r's display is their instability. For example, consider the situation where the true proportions of variance for G, E, and I are .70, .20, and .10 respectively. Then using simulated twin samples which tend to even be a bit larger than those used by Jensen, we see that the standard error of estimate of the interaction coefficient is 0.0915, or nearly as large as the coefficient itself. It is thus *not unlikely* that if the "true" interaction component was .10, Jensen would have found its value to be approximately zero, given the data available to him.

Suppose we even take a more extreme example in Jensen's favor, as his strongest argument appears to be that the heritability component is high, and approximately .75. If we let the components of variance for G, E, and I be .80, .10, and .10 respectively, then the standard error of estimate for the interaction component is still very high at 0.0895. Once again, given his sample size, it is not unlikely that Jensen would find a negligible interaction component when in fact the true component was 0.10. Yet, we have discussed earlier the great importance of

even a small interaction component in accounting for racial differences in observed IQ.

Finally, if we compare the standard errors of the median estimates given in Table 4 with those which would have been attendant to a *single large twin study, we see that the standard errors from the median estimates are approximately equivalent to having a single large study of 100 twins.* Thus, the estimates of the model parameters which Jensen uses as the underpinning for his entire analysis have approximately the same reliability as if he had selected one sample of 100 twins. That Professor Jensen would draw such far-reaching implications from a sample of this size, and in the complete absence of any black twin studies data at all, is remarkable.[10]

Quality of the Data:

Not only is the statistical reliability of Jensen's parameter estimates fairly low, but the quality of the twin studies data may be questioned as well. Jensen himself is well aware of the problems which attend testing young children from less advantaged backgrounds.[11]

When I worked in a psychological clinic, I had to give individual intelligence tests to a variety of children, a good many of whom came from an impoverished background. Usually I felt these children were really brighter than their IQ would indicate. They often appeared inhibited in their responsiveness in the testing situation on their first visit to my

[10] Our simulations averaged 225 pairs of twins per replication of [illegible] studies, and produced estimates with an average variance that would have been found in a single large study with 145 pairs of twins. The efficiency of the median is thus about 145/225, or 64.4 percent. Professor Jensen, for three MZ studies, reports 114 pairs of twins, and if we assume the fourth study had 40 twins, the effective number of pairs is about 100. (64.4 percent of 154 is 100.) The proper procedure for pooling intraclass correlations, drawn from the same population is, of course, by the use of Fisher's Z transformation, and not by taking medians. For the reader who wonders how seemingly minor technical details could modify the broad outlines of Professor Jensen's original arguments, we can only point out again that a few percent of variance can be responsible for large differences between group means. Yet we are dealing with statistical procedures which, even when perfectly applied, are subject to large sampling errors. To ignore such technical details is to ignore the crux of the argument.

[11] The last sentence in Jensen's quote is perhaps worth a note. If we wanted to measure the purely genetic component of IQ and to ignore or avoid the environmental component, then we would certainly not view the 8 to 10 point gain as having "much of anything to do with changes of ability." Clearly, if we are looking only for genetic factors, then all the environmental component is measurement error. But when we are attempting to resolve the apportionment controversy, we must, as Jensen said, rely solely on an operational definition which equates ability to the IQ score. If the score changed, ability changed. Jensen's inclinations in this matter are surprising.

office, and when this was the case I usually had them come in on two to four different days for half hour sessions with me in a "play therapy" room, in which we did nothing more than get better acquainted by playing ball, using finger paints, drawing on the blackboard, making things out of clay, and so forth. As soon as the child seemed to be completely at home in this setting, I would retest him on a parallel form of the Stanford-Binet. A boost in IQ of 8 to 10 points or so was the rule; it rarely failed, but neither was the gain very often much above this. So I am inclined to doubt that IQ gains up to this amount in young disadvantaged children have much of anything to do with changes in ability. (p. 100)

In other words, Jensen feels that certain prior environments 'inhibit responsiveness' and thereby lower IQ scores rather uniformly, by about 8 to 10 points—which is more than half the difference between the white and black mean IQs found by Coleman. Because of our national prejudice against the black person, we would expect that the black child would learn quite early in life that whenever he was in a setting in which 'evaluation' was likely to occur, then the experience was probably going to prove unpleasant. He may thus display inhibited responsiveness, possibly of a kind not so easily overcome by fingerpainting.[18]

Internal Consistency of Parameter Estimates:[19]

1. Monozygotic Twin Studies:

The monozygotic twin studies summarized by Professor Jensen estimate the genetic component of phenotypic variance only if both members of the pairs are identical genetically and otherwise assigned to environments at random. Besides their genetic endowment there is, after all, one other thing identical twins have in common: they presumably spent the first nine months of life in identical

[18] There may be some confusion between environmental effects and unreliability of IQ scores. Think of a good wall thermometer; the small variations in its reading from occasion to occasion are not unreliability—they are the changes in the room temperature, and it was for the purpose of detecting them that we put the thermometer in the room. To assume that any fluctuation in measured ability subsequent to conception must be due to unreliability in the measurements is to fall into the concept of fixed intelligence which Hunt so completely demolished. Of course, such environmentally derived gains or losses in IQ are not maintained unless the environmental factors which contributed to them also persist.

[19] The reader will need to have one point firmly in mind throughout the following discussion. All three kinds of pair studies—MZ twins reared apart, MZ twins reared together, and foster children reared together—estimate two proportions. Each directly estimates the proportion of phenotypic variance due to whatever factors both members of a pair have in common. Unity minus that number estimates the proportion of variance due to whatever factors are not identical for both members of a pair. The procedure is completely blind; it does not specifically estimate genetic factors or environmental factors. Whatever affects both children in each pair identically, appears in the direct estimate. Whatever has been randomly assigned to both children in a pair, appears in the indirect estimate.

environments. Jensen argues at length that the prenatal environment may carry a large part of the total environmental impact upon IQs:

Thus, much of the average difference between MZ twins, whether reared together or reared apart, seems to be due to prenatal environment factors. The real importance of these findings, of course, lies in their implications for the possible role of prenatal environment in the development of all children. It is not unlikely that there are individual maternal differences in the adequacy of the prenatal environment. (p. 68)

And, he later adds:

There is no doubt about the fact of the greater prevalence in poverty areas of conditions unfavorable to optimal pregnancy and safe delivery. (p. 69)

But those differences are included in the 75 percent of the phenotypic variation assigned to genotypic allotments, merely because the MZ twins have them in common. Further, the foster pair studies would also *underestimate* the environmental variance for the same reason. The foster children did not have common prenatal environments.

Jensen further indicates that:

In pairs of identical twins, the twin with the lower birth weight usually has the lower IQ (by 5 to 7 points on the average) at school age. This is true in both white and Negro twins. The birth-weight differences are reflected in all 11 subsets of the Wechsler Intelligence Scale for Children and are slightly greater on the Performance than on the Verbal Tests (Willerman and Churchill, 1967). The investigators interpret these findings as suggesting that nutrient supplies may be inadequate for proper body and brain development in twin pregnancies, and that the unequal sharing of nutrients and space stunts one twin more than its mate. (p. 68)

Since this is a *difference* between the twins, it will show up in the part of the total variance not accounted for in the genetic intraclass correlation. It has been added to the ($1.00 - .75 = .25$) environmental column. But the great majority of the natural population are singletons, not twins, and they would not be subjected to an environment which included competition from their womb-mates. Therefore, we may be misestimating the total variance in the natural population when we make use of the MZ twin studies, and also *overestimating* the effects of environment.

We do not know how to resolve this potential conflict in estimation; the MZ twin studies might have underestimated or overestimated P_E. Our point is merely that the effects may or may not cancel, and thus the estimate of P_E may or may not be biased.

2. The Foster-Pair Discrepancy:

Jensen points out:

> Another interesting comparison is between MZ twins reared together ($r = .87$) and reared apart ($r = .75$). If $1.00 - .75 = .25$ (from the MZ twins reared apart) estimates the total environmental variance, then $1.00 - .87 = .13$ (from MZ twins reared together) is an estimate of the environmental variance *within families* in which children are reared together. Thus, the difference between $.25 - .13 = .12$ is an estimate of the environmental variance *between families.* (p. 51)

That would certainly be the case. Except that the estimate of the proportion of total variance due to environmental differences, which is derived from studies of unrelated foster pairs raised together, is also an estimate of only the *between family* differences in environmental conditions. That estimate, from Jensen's Table 2, was .24, not .12. In fact, if we add the several estimates up:

Genetic factors:	
(from MZ twins reared apart)	.75
Within family differences	
(from MZ twins reared together)	.13
Between family differences	
(from unrelated foster pairs raised together)	.24
	1.12

We have just accounted for 112 percent of the total variance, which is not possible. Thus, there is some question as to how to appropriately adjust these three components so that they add correctly.

3. Covariance Between Heredity and Environment:

Jensen states that:

> Children with better than average genetic endowment for intelligence have a greater than chance likelihood of having parents of better than average intelligence who are capable of providing environmental advantages that foster intellectual development. (p. 38)

In other words, Jensen implies that if the covariance term between heredity and environment exists, then it should be positive.

Suppose that in accordance with Jensen's suggestion, we assume that there is a positive covariance term in the 'real world' sampled by the studies of MZ twins reared together, but that this term also appeared in both the MZ twins reared apart and in the foster pair studies. While this would be experimental error, one

might suspect that neither twins nor foster children are assigned to substitute families at random. This approach might help us to "repartition" the components of variance so that they add to 1.00. In other words, suppose we assume the covariance term was 'counted twice;' it was included in the estimate of .75 attributed to genetic factors by the MZ reared apart studies, and then included again in the .24 attributed to between families environmental conditions by the foster pair studies. Let's see what happens:

P_G is the proportion of variance due to genetic factors
P_B is the proportion of variance due to between family environmental differences
P_W is the proportion of variance due to within family differences
P_C is the proportion of variance due to the covariance of genetic and between-family differences

We know that all four proportions must add to 1.00 by definition. We also have data from three types of studies (and we are assuming that the total variance was the same in all three sets of studies), which gives us three additional equations:

By definition:
$$P_G + P_B + P_C + P_W = 1.00$$
From the MZ twins reared together studies:
$$P_G + P_B + P_C = .87$$
From the MZ twins reared apart studies:
$$P_G + P_C = .75$$
From the foster pair studies:
$$P_B + P_C = .24$$
We can then solve for the following estimates:
$$P_G = .63$$
$$P_B = .12$$
$$P_W = .13$$
$$P_C = .12$$

The heritability has been reduced to .63 and the environmental component increased to .37. Although we do not wish to argue that these revised proportions are in reality correct, this again points out that Professor Jensen's analysis contains an inconsistency which must be resolved.

4. Combining the Unreliability and Foster Pair Discrepancy Arguments:

If we combine the results of our twin study simulations from Table 4 with the inconsistency argument presented above, we see that there really need be no discrepancy at all, and that the four components may not sum to 1.00 simply because of sampling errors. Our simulations have shown that estimates of the median intraclass correlation from five fairly small studies are quite erratic when the true proportion of variance due to the factor equated within pairs is small. Thus, perhaps the true proportion of variance between families is in reality only .12, and the median value of .24 came about due to sampling errors. (In fact, given Jensen's sample sizes, the standard deviation of the distribution of medians is about .08 when the true proportion of variance is .12. Thus, using a normal approximation, the observed median value of .24 is *not* significantly different from a hypothesized value of .12 at the .05 level of significance, one-sided or two-sided.) However, with that kind of error lurking in the wings, the stage machinery of a Model II ANOVA may not be the best way to dramatize the features of our world's environments which have raised or afflicted our children's intellectual capacities.

Summing Up:

In concluding this section, then, there are serious questions which need answering before Jensen's parameter estimates can be accepted. We have shown how a small amount of interaction between genetic endowment and environment can easily explain how two races with identical genetic endowments can have large differences in mean IQs. We have further argued that even assuming the twin studies data were of excellent quality, the data are not at all statistically inconsistent with the existence of something like a .10 interaction component of variance, and that this magnitude of interaction could account for mean differences of more than one standard deviation in black-white IQs without any genetic differences between races (see Table 3). Finally, the inconsistencies of the data in their additive behavior, together with the complete lack of any data at all on black twins, in our judgment wrecks the credibility of even a tentative assertion of genetic differences in intelligence between races.

Statistical Assumptions and Model Considerations:

As the study of race and intelligence is invariably shrouded by both political and statistical debate, we wish to take the liberty of offering a few observations on the inherent problems of measurement and model selection.

Models of Intelligence

RICHARD J. LIGHT AND PAUL V. SMITH

All scientists know that one of the most crucial stages in the research process is the model selection. For example, Jensen has selected the Model II ANOVA to estimate his parameters, and therefore implicitly agrees to be bound by its assumptions, limitations, and descriptive ability of the real world. After all, incorrect assumptions will permit meaningless or misleading conclusions to be drawn from extensive, perfect information. Thus, the selection of a model has the limited virtue of making obvious its dependence upon the worth of the direct measurements it requires.

If this comment suggests that we believe that scientists cannot really resolve controversies such as this until direct identification of genotypes and environments is possible, that is correct. Jensen says:

> Determining the heritability of a characteristic does not at all depend upon a knowledge of its physical, biochemical, or physiological basis or of the precise mechanisms through which the characteristic is modified by environment. Knowledge of these factors is, of course, important in its own right, but we need not have such knowledge to establish the genetic basis of the characteristic. (p. 44)

We believe Jensen is incorrect: determining the heritability of a characteristic in a single population not only depends upon a knowledge of its physical, biochemical, and physiological basis but also on an equally exact knowledge of its environmental contributors. Where we have reason to believe that we can identify and isolate the causes of a characteristic experimentally, we can control for or randomize various combinations of factors and thus can employ certain statistical techniques with some immunity from gross error. Without such *a priori* knowledge, we would be driven to consider land-ownership during the middle ages to be a sex-linked dominant trait in first-born males with a coefficient of heritability near unity. Our knowledge that the currently proposed genetic mechanisms cannot single out the first-born prevents us from drawing that conclusion, and substantial documentation about the social system at that time corroborates our decision.

However, when someone asks whether the difference in observed mean IQs between white and black populations is attributable to genetic differentiation, a fundamentally different problem is being posed. It bears directly on the choice of a components of variance model to investigate racial differences.

Any components of variance model presupposes that the sources of variation can be distinguished and crossed in the experimental situation—*not only can be, but were*. Placing a covariance term in a model handles the possibility that two

sources of variation are linked in the real world—but only conceptually. Every study Professor Jensen reports has neglected the most obvious linkage between genetic and environmental factors: that the environment is composed of sentient beings who note genetic differences and act to differentiate them further.

A genetic characteristic revealed by external measurement can function both as a sign and as a signal. For example, when we detect the blueness of an eye we have the sign of the presence of a specific genotype. The blueness of an eye may also be the signal for a set of social processes to occur (love, or so we are told); but that entails no difficulty in deriving estimates of heritability coefficients, because we have quite definite information to the effect that environmental conditions cannot change eye color. But if the environment can also affect the measured characteristic, then the possibility of a feedback loop is real: that is, the environment reads the sign of the genetic variation as a signal for initiating social processes which enhance the variation; the enhanced variation in the characteristic is then again taken as the signal for another round of differentiation, and so forth. Further, we have no reason to believe that the enhancement of individual differences would be either as strong or as accurate for one social group as for another.

We certainly know that human environments do read genetic signs as social signals, and not just for romantic love: a darkened skin is more than the genetic sign of the transmissable presence of melanine—it has become the signal for an array of social processes in our society whose scope is currently unknown and whose nature is perhaps best explicated by others. The essential point, however, is that monozygotic twins reared apart have not been stripped of the signs of genetic differences in IQ that could signal the occurrences of other social events. Jensen, in dismissing Rosenthal's experiment on teacher expectations on statistical grounds, failed to note that it is one of the very few instances in the literature when a researcher deliberately tried to mislead the signal-seeking component of a social system about a genetic sign. Our ability to account for a difference in population means is wholly dependent upon the success of such separations as the one Rosenthal attempted.

Serious future research into questions of race, environment, and genetic endowment must take into account this environmental feedback process, both in the choice of a model, and in the design of surveys or experiments to collect data. Otherwise, many of the crucial processes which affect IQ will be ignored, and research results may be little more than projections of the assumptions of their authors.

Environment: The Cumulation of Effects Is Yet to be Understood

ARTHUR L. STINCHCOMBE
University of California, Berkeley

Professor Stinchcombe deals with the Jensen article from the point of view of an "environmentalist" but not from the simplistic stance that Professor Jensen attacked in his original article. Stinchcombe argues that deprivation does more than prevent children from learning simple skills at an early age—that cultures or social conditions must operate consistently and sequentially to produce successively higher levels of cognitive functioning. Environments, he argues, are cumulative, and until researchers can account for the complexity of environment, statements about the proportional effects of heredity and environment are premature. Thus extrapolations from twin studies limited to a single social group to estimates of the genetic capabilities of a different group are particularly suspect.

This essay is divided into two broad parts. First I want to make several comments about what Jensen's evidence shows, taking the evidence at face value. In general, Jensen conceives the evidence he presents to have a single obvious interpretation. I think the singularity of interpretation is often the result of a kind of systematic naivete about how the environment works.

Second, I want to develop the outlines of a theory about how the environment determines the abstract cognitive structures people have to deal with the environ-

Harvard Educational Review Vol. 39 No. 3 Summer 1969

ment—especially the symbolic environment—which are measured by IQ tests. The basic argument will be that many of the findings in Jensen's article fit into a pattern that would be produced if rural and oppressed cultures were inefficient in teaching advanced cognitive structures at later ages. Jensen assumes that if environment works harm, it will do so in the learning of simple cognitive skills at earlier ages. This might be called the "Headstart" environmental theory. When he gives arguments against this theory, he thinks he is arguing for a genetic theory. But he is also implicitly arguing for a "civilization" theory of environmental determination of thought, which says that cultures differ most at the most sophisticated levels. My purpose will be to use his evidence in favor of a different theory of environmental impact on IQ.

What Do the Kinship Data Show?

The most impressive argument for the heritability of intelligence is clearly the data on correlations among kinsmen in IQ. The basic finding is that if you take pairs of kinsmen that are genetically more closely related (*e.g.*, identical twins) and pairs of kinsmen that are farther apart (*e.g.*, fraternal twins), then the correlation between the IQs of the closer pairs will be greater than the correlation between the IQs of the more distant pairs. That is, if you know the IQ of one identical twin, you can predict the IQ of the other very well. If you know the IQ of a fraternal twin, you cannot predict the IQ of the other nearly as well. By combining genetic theory about *differences in genetic closeness* of different types of relatives with knowledge of the *differences in correlations* between pairs of different types of relatives, one can estimate how much effect genetic closeness has. From this estimate of the effect of genetic closeness, one can estimate the *size* of the effect of genetic constitution, even though one cannot measure genetic constitution itself. This seems to me a valid procedure for what it is designed for, though subject to the risks of any attempt to get at unobserved variables from the pattern of manifest variables.[1]

But if we look at Jensen's estimates of environmental effects for IQ (p. 51), we find that the difference *in environment* between one twin and another is the same size as the difference *in environment* between one family and another. That is, if

[1] One assumption it does make is that if two people are born in the same family they have the same father. This is, of course, a theory of the frequency of extramarital affairs or of extramarital contraception. It has been estimated that a fourth of all married women in the U.S. have extra-marital affairs, but I know of no studies of their contraceptive practice in such affairs. One could estimate the frequency of such fruitful affairs by studying some clearly inherited characteristics, such as blood types.

Environment
ARTHUR L. STINCHCOMBE

you pick out two brothers at random, the chance that one of them will be in a "deprived environment" while the other is in an "advantaged environment" is the same as the chance that one out of a pair of randomly picked families is a "deprived environment" while the other is advantaged. This clearly is not the notion most people have of variations in environment.[2]

This peculiar fact of as much environmental variation within families as between families is illuminated if we look at the samples. The one described for London is more or less typical. To summarize briefly what is going on, if you and your cousin both live in London and both are still in school, then the difference between your father's family and your uncle's family is about of the same order of magnitude as the difference between your family as you grew up in it and your family as your brother grew up in it. That is, the environmental variations we are talking about are variations *within* an endogamous group, a group that intermarry and exchange children by adoption. If we want to generalize across endogamous groups (groups that intermarry among themselves), for instance across races, we have to ask whether there is more variation in environment between a white and Negro family than there is among kinfolk living in London. If there is, then the between-families variance in environmental factors would increase, and the relative role of genetic factors would decrease. Also, the relative role of within-family variation in environment would decrease relative to between-family variation.

What is really needed to apply such kinship research to the problem of racial differences are pairs of identical twins, one of whom is a Negro, and pairs of fraternal twins, one of whom is a Negro.

Regression Effects

Another apparently impressive piece of evidence is the larger regression effect for Negroes rather than whites (pp. 83-84). Briefly the data are mean IQs of children, subtracted from the mean IQs of parents in each of three, four, or five social classes. (Roughly speaking this is what is going on—the actual calculations

[2] Incidentally, it may confuse some readers of Jensen's article that he takes correlation coefficients as measures of variance explained at this point in his argument. He does not explain why, and most people learned to treat correlations as the square root of variance explained. But Jensen is right here. The key is that the estimate of the correlation between the *unmeasured* genetic variable and the manifest variable is estimated by the square root of the correlation between two manifest variables assumed to have the same correlation with the unmeasured variable. When this square root is then squared to give the variance explained, it gives back the manifest correlation. In psychometric language, the reliability of a test is an estimate of the square of its validity, *i.e.*, the "variance explained" of the test by the variable it measures.

do not seem to be as adequate to the purpose as I described.) The differences are larger for Negro middle class children than for white middle class kids. The argument here goes as follows. The difference between childrens' and parents' IQ measures the difference between the particular set of genes that are manifested in the parent and the average of the total set of genes which the parent carries. That is, out of the total pool of genes that the parent carries only one subset are the ones that he gives his children. If the manifested subset is very different from the pool, then the children will be very different from their parents. So the size of the difference between parents and children is supposed to measure the distance *for the parents* between the genes they manifest and the genes they carry without manifesting. If Negroes of high intelligence show more regression effect than whites of the same intelligence and environment, that is supposed to show that the pool of genes from which exceptionally talented Negroes come is farther from them, on the average, than the pool from which exceptionally talented whites come.

Now the key to this inference is actually equating the environments of the children. For if the children of talented Negroes are actually exposed to a less cognitively rich environment than are whites, even though they are in the "same" social class as measured, then we would expect a larger regression effect even if the pool of genes were the same distance from the parents. What we have to do then is to examine the equation of parents' status first, and then whether the equation of status equates environments. Briefly the argument will be that Jensen has not in fact equated statuses, and that even if he had he would not have equated environments.

First, let us assume that the indicator of status in the studies that Jensen quotes were a perfect measure of status. Then what he has done is to pick out three (or four, or five) cutting points on a continuous variable, and equated all people above the top cutting point as middle class, and so forth. Now as Jensen so carefully shows for IQ, if one distribution has a lower mean than the other and the distributions have the same shape, the more extreme values are more underrepresented on the lower distribution than the more moderate values. The same is true of status. Rockefellers, Mellons, and du Ponts are more underrepresented among Negroes than are $20,000 a year men; $20,000 a year men are more underrepresented than $10,000 a year men, and so forth. So if you take a given cutoff point for middle class, those few Negroes that are middle class will be mostly very near the cutoff. Many of the whites will be quite far from the cutoff. Thus actually the mean class position for middle class Negroes is likely to be much lower

than the mean class position of middle class whites. Thus even if the class measures were perfect, the groups would not be equated.

But of course the actual situation is that the measurement of social status is much more unreliable than the measurement of IQ, and Jensen is always careful to take account of measurement error for IQ. Merely to take occupation as an example, at a given point in time when the census reinterviews people about their occupations, they get 17 to 22 per cent giving a different occupation (Blau and Duncan, p. 15). Of course the census uses trained interviewers who know what is required to classify an occupation. Most school studies ask children to describe their parents' occupation. Census studies of the reliability of occupational scales recently reported give figures in the eighties, somewhat lower for Negroes than for whites. Questionnaires, especially from children, must have a great deal more measurement error than census interviewer

Furthermore, the environment of a child is a cumulative matter, consisting of so and so many months in a middle class family when his father had a good job, so and so many in a welfare family when he was unemployed, and so forth. Thus even if we had a perfect measure of the status of a family at a given time, it would be a poor measure of the cumulative environment of the child if peoples' status changes much over their lives.

Perhaps this will be more understandable with an example. Suppose one truly has 10 per cent of middle class Negroes and 90 per cent working class, while whites are 50 per cent middle class and 50 per cent working class. And, also, assume that the likelihood of misclassification is the same for both races and both social classes. Now apply a measurement that misclassifies 10 per cent of all people. Applying this to Negroes, we get

	Classified as Middle Class
90% of the 10% Truly Middle Class =	9%
10% of the 90% Truly Working Class =	9%
Total classified as Middle Class =	18%

Percentage of Negroes Measured Middle Class Who are Truly Workers = 50%

Applying the same measurements to whites, we would get

90% of the 50% Truly Middle Class =	45%
10% of the 50% Truly Working Class =	5%
Total Classified as Middle Class =	50%

Percentage of Whites Measured Middle Class who are Truly Workers = 10%

Thus we would expect that many more of the Negroes Jensen calls middle class are truly working class (either at present, or in the cumulative environmental sense) than are the whites. Hence because of measurement error he has not equated the environments. Actually in most of the studies he reports on, the most important environmental variable, the IQ of the parent, has not been equated at all.

Finally, of course, children associate with other children as well as with their parents. With a highly segregated society, this means that Negro children associate with workers' children, while white children associate with middle class children. Consequently even were the status of families equated, the status of interpersonal environments of children would not be equated.

On all these grounds, then, the different regression effects cannot be taken as evidence that talented Negroes are farther genetically from their gene pools than are talented whites.

How Big is a Big Environmental Effect?

The difficulty Jensen has in thinking about environment as cumulative shows up again in his evaluation of the size of environmental effects. Jensen would like to have environments change once and for all from one environment to another, as genes do, and then stay that way. We might call this the spurt conception of the environment. Environmental change in this conception comes in a spurt, and any time after that we can measure the effects of the spurt. Genetic variation does come in spurts; environmental variation does not.

The difference between Negro and white measured IQ scores in the U.S. is about one standard deviation, or 15 IQ points. Jensen observes that by decreasing the fear people have in the testing situation, so that they feel they control the situation and can keep from getting hurt, one can often increase poor children's IQs by 5 to 8 points, or about one-third to one-half of the distance between whites and Negroes. Now this seems to him to be evidence of a small environmental effect, since most *other* changes in environments only produce about the same effect. But this is only because he regards taking fear out of social situations as eliminating measurement error. If he regarded it as changing the environment, he would conclude that taking fear out of the relations of Negroes to their environment might, by itself, decrease the difference between Negro and white performance.

Then Jensen goes through a set of studies which show that by changing the environment of children for six, or nine, or three months, one can change their IQs by 5 to 10 points. That is, a short period of environmental enrichment can

Environment
ARTHUR L. STINCHCOMBE

apparently wipe out a third of the difference between whites and Negroes. Furthermore, this is change in only 25 or 30 hours a week of their environment.

Finally, we quite often find that naturally occurring variations in environments cause variations of this order of magnitude in IQ or mental ability scores. Jensen himself quotes a study which estimates that a standard deviation of environmental variation within a white group can cause two-fifths of a standard deviation in IQ. To explain the variation between Negro and white IQ on this enironmental basis, we need to find determinants of IQ on which there is a difference of two and one-half standard deviations between Negroes and whites, assuming a linear relationship between the variables. This is a difference such that, for a normally distributed variable, about one per cent of Negroes would be above the average of the whites. It does not seem to me outrageous that a well-measured variable of oppressiveness of conditions of life and cultural deprivation might show such a difference.

Besides this estimate of Jensen's, we can get a notion of the size of the environmental variation by observing variation between Negroes whose environment approximates that of whites from those whose environment approximates that of Negroes. Clearly if Negroes go to nearly all-white schools, they are exposed for 30 hours a week to the same environment most whites are. If they go to all Negro schools, the 30 hours a week are like those to which Negroes are exposed. James S. Coleman's estimates of the correlation between proportion white and Negro mental abilities indicates that by changing this part of a child's environment one makes about a half a standard deviation difference. This is of course also a non-cumulated score. A Negro child in an all-white school is much more likely than a white child in the same school to have spent part of his life in an all-Negro school.

The estimate of the effect of variations among white families in the kinship studies comes to approximately the same conclusion. If the correlation between adopted children and their parents in IQ is taken as a measure of the effect of environment, environmental variation between families who adopt each others' children explains about 24 per cent of the variance. That means that a standard deviation of environmental variation in families that adopt each others' children would cause about a half standard deviation variation of IQ.

What we have then is a large number of environmental variables, many of them occupying a small portion of children's lives, that explain between one-third and two-thirds of the difference between races. Equalize the amount of fear between races, put Negroes in all-white schools, or schools with the same conditions, improve the standards of Negro families by a standard deviation, then

(if the effects are additive) the IQs would be equalized. Jensen thinks that each such variation is small because it only explains half or two-thirds of the racial difference. But environments cumulate.

Because Jensen thinks of environment in spurts rather than in hours per day, he is "surprised" when an equalizing spurt does not equalize IQ. He translates this surprise into an argument that the spurt that created the genetic differences is more important. Suppose that we have a difference in environment which gives one group a rate of improvement in cognitive skills of 5 per cent a year, while another group advances at 3 per cent a year. The ratio of advance would be 1.05/1.03, or about 1.02 per year. If the two groups start equal, the more advantaged group will be about two per cent ahead after a year. If this environmental difference is maintained for, say, the first 20 years of life, the cumulated effect is about $(1.02)^{20}$, or about 1.5. The effect of a given year's enriched environment, which equalizes the increase for one year, is about a fifth of the disadvantage at 5 years of age. But it is about one-twenty-fifth of the disadvantage at 20 years of age. Jensen seems surprised that equalizing a single year results in an effect which disappears in a few more years. He would not be so surprised if he had thought a bit about a compound interest table.

The Cognitive Environment of the Individual

So far I have let Jensen get away with his conception of intelligence, and of the environment, because most of his arguments can be undermined even if we allow him his conception of his own business, and require only elementary elaboration of his conception and measurement of environment. It seems to me that the rest of his argument stands up at its own level, and requires some more fundamental thought about cognition and the sociology of cognition. What I will try to do now is to develop a theory of what intelligence, as measured by the usual IQ tests, is, and then on this basis develop a notion of what kind of environmental variables might determine it. Then I will try to derive the remainder of Jensen's results from this alternative theory. The combination of his genetic spurt conception of intelligence and his Headstart theory of environment makes it hard for him to think about cognitive functioning appropriately.

First, to intelligence. The central fact about intelligence among children is that older children have to be able to do more than younger children in order to stay at the same level of intelligence. Jensen notes that Negro children's capacities go down from a slight superiority to whites at early ages to inferiority at later ages. That does not mean, of course, that Negro children forget by five years old

what they knew how to do at three. It means instead that the time between mastery of one task or test and mastery of a more complex task or test is longer for Negro children than for white children.

That is, from the point of view of an individual child's biography, his intelligence is a measure of the time between being able to manage one type of task and being able to manage the next. Thus tasks are arranged more or less in layers with one group of tasks or intelligence test items being a layer which is normally learned after the "simpler" ones and before the more "complex" ones.

What seems to be behind each layer of tasks is one or more *cognitive structures*, or styles of abstraction and reasoning. Once a child learns to manage a given general cognitive structure or style of abstraction and reasoning, he can usually solve or answer most of the tasks or test items using that cognitive structure. The cognitive structure may involve perception, operations on perceptions, combinations of perceptions with previous knowledge, and so forth. There is a fair amount of evidence that such structures are well ordered, so that a five year old average IQ score is rarely made up of a hash of questions at a four year old level and a fifteen year old level. Instead the normal five year old answers four year old and five year old questions. The six year old answers 4, 5, and 6 year old, etc. The classic work here was done by Jean Piaget. The funny scatter diagram at p. 114 in Jensen seems to be evidence in Piaget's favor. (The subtitle of these diagrams is mixed up with something else). If we think of intelligence as such a layer-cake type of phenomenon in the individual, rather than as a manifestation at the present time of his constant IQ "dispositional property," then we can ask what sorts of environmental features would be most likely to affect it. Probably the most important characteristic of the environment is the frequency with which the *next developmental type* of cognitive structure is used in a child's environment. Most of us at breakfast use grunts and groans and expressions of displeasure whose cognitive content is easily mastered, if not enjoyed, by a three year old. During the day we vary in intellectual level, returning just before sleep to the same level of grunts and groans, if we are lucky. The same variation is true of a child's playmates—they mostly function below their capacities, and only occasionally reach their limits. Although Jensen sometimes seems to subscribe implicitly to the Education School Fallacy (that anyone is smart enough to teach kids something),[3] the same variation in intellectual level is characteristic of classrooms.

[3] Jensen, for instance, wonders why experimental programs, when made mass programs, very often have less effect. Perhaps the mass of teachers are dumber than those involved in experimental programs.

What we would like, then, as an environmental measure of development pressure, is a measure of the density of interaction of the child with someone who, in the interaction, leads him up one cognitive structural level from where he is, solves his problem, then restates it at the level where the child is (Compare the account of language learning in Roger Brown, *Social Psychology* pp. 292-297). Since most people most of the time function with structures much less complex than they manage in their finest hour, a person's cognitive behavior is probably distributed with different intensities at different levels below his highest capacities. These intensities or densities are probably determined by various influences, such as the intelligence of people around him, the complexity of the problems in his work, his amount of education, and so forth, as well as being limited by his upper limit of cognitive structures (his IQ). It is well known, for example, that professional men's IQs do not decline with age as rapidly as do the IQs of manual workers. Presumably this is because they keep the highest levels of mental functioning in better shape by more practice.

Thus there should be two main causal forces which would determine the mean level and variations of the cognitive functioning of a man—his *capacity to abstract* (his IQ) and his socially and psychologically induced *inclination to abstract* (determined by his education, his occupation, the intelligence of his wife and friends, and their educations and occupations, etc.). Thus we would expect to find correlations between people's IQ and the rate of development of children with whom they come in contact. For instance, there should be correlations between the IQ of foster children and the IQ of their foster parents (reported by Jensen, p. 52), between the IQ of teachers and the achievement of children in their schools (reported by Coleman pp. 317-318—the effect is stronger at higher grades), between the IQ of peers and the rate of advance of the IQ of peers (difficult to isolate, but fairly clear in some Coleman data, e.g. at p. 307).

Given the proportion of a person's cognitive functioning which just strains the child, the next question is the *hours of attention* he gives to the child, and the *attentiveness of the child* during those hours. We have very few estimates of these things, but we suppose that mothers spend more time with children than fathers (hence there should be higher correlations with mothers IQ), that parents spend more time with only children than with many children (hence only children should be smarter generally and have less regression from parents' IQ), and so forth. The hours of attention received by and paid by children at different levels of cognitive tasks are ordinarily not measured in educational experiments. If the change from small to large programs changes the level of atten-

tiveness of teachers, or students, or both, then we would expect less change in cognitive functioning in the expanded programs.

Civilization and Intelligence

We could now define the level of civilization of a group by the average level of cognitive function used to solve its problems. That is, an "advanced" civilization would be one whose average man at the average moment was using, say, a third grade level of abstraction. There would be a wide distribution around that level, of course. A "backward" people would be one which did not ordinarily use the cognitive structures (e.g. mathematical reasoning, experiments, etc.) of which it is genetically capable. In general, more complex cognitive processes require reading and writing. Few people can solve mathematical problems in their heads that they solve easily on paper. Probably a good surrogate indicator of the level of cognitive functioning of a group is hours per day spent in reading and writing.

Ever since a famous sociologist of cognitive affairs in 1848 spoke of "the idiocy of rural life," evidence has been accumulating that farm life has not provided the practice in symbolic analysis of the environment that a city did. Civilizations grow in cities. In some rich agricultural areas (e.g. Iowa, New Zealand) civilization has more or less completely spread to the countryside.

If we suppose that the definition given above of civilization is something like the same thing that we ordinarily call civilization, then it will come as no surprise to find that rural people when tested have regularly turned up less intelligent on IQ tests. Furthermore groups imported for plantation labor and not allowed to develop their own civilized institutions (schools, churches, publishing houses, newspapers, political parties, trade unions, welfare groups, community governments, etc.), show the most IQ disadvantage compared to urban people. They also show the most gains in IQ when moved to the more civilized environment of an urban slum. They show still further gains as the hours per day of reading and writing increase in the group.

What this suggests then is that one of the legacies of slavery in the Negro community is a dearth of institutions that function routinely at advanced cognitive levels. There are too few Negro colleges, too few professionals, newspapers and book publishers, and so forth, to fill the environment with a rich variety of cognitive levels. But it might be that as Negro children grow older, after starting to learn with the same capacity as white children, their environment is progressively less rich in the frequency of use of the next developmental cognitive structure.

If this were the case, we would expect that Negro IQ would appear to decline,

77-942 O - 72 - 48

as it took them longer to learn each successive cognitive structure. This is what Jensen reports (e.g. pp. 86-87; p. 77 gives the same for social class). We might also expect there to be stability of IQ disadvantage, even though this looks as if it might be evidence for genetic determination of IQ.[4]

The purpose of these last two sections has been to develop a more serious environmental theory of intelligence than the Headstart theory, to suggest how research might be done on such a theory, and to explain why we might expect on the basis of such a theory that an oppressed rural group from a backward section of the country might show a progressive deficit. Instead of defining environment in a way borrowed from other fields, an environmental theory of determinants of cognitive capacity needs to define the environment in intellectual terms.

In particular, we need to know the proportion of time a child focusses attention on cognitive structures at the next level of development above his own. Various results that we already have can be interpreted in the light of such a conception of the environment. For instance, the correlation of children's achievment with the mean vocabulary scores of their teachers makes sense if we think that smarter teachers will strain their students' capacities a little more.

The idiocy of rural life, the cognitive consequences of growing up in urban slums, the disparities of achievement among ethnic, religious, and social-class groups begin to make sense if we define civilization in terms of the densities or frequencies of use of intellectual structures at different levels.

[4] An I.Q. deficit of 15 points, or one standard deviation, is approximately equivalent to the amount of development that an average child has gone through in the last sixth of his life. That is, for a child of six, an I.Q. of 85 means he can do approximately the things a normal 5 year old can. For a child of 15, it means he can do approximately the things a normal 12 year old can. Thus a constant deficit of one standard deviation means functioning at progressively more years behind others.

References

Blau, Peter M. and Duncan, Otis Dudley. *The American Occupational Structure*. New York: John Wiley and Sons, 1967.

Brown, Roger, *Social Psychology*. New York: The Free Press, 1965.

Coleman, James S., and Others. *Equality of Educational Opportunity*. Washington: U.S. Government Printing Office, 1966.

Jensen, Arthur R. How Much Can We Boost IQ and Scholastic Achievement? *Harvard Educational Review*, 39, (Winter 1969), 1-123.

*Happenings on the Way Back to the Forum: Social Science, IQ, and Race Differences Revisited**

MARTIN DEUTSCH

New York University

Professor Deutsch reviews the literature on compensatory education, intelligence testing, and the nature of educational environments and concludes that Arthur Jensen had constructed an article which had "negative implications for the struggle against racism and for improvement of the educational system." Deutsch believes the Jensen article holds a consistent bias toward an undemocratic eugenic and racist hypothesis.

The publication in the Winter issue of this journal of the long article by Arthur Jensen ("How much can we boost IQ and scholastic achievement?") has resulted

* I should like to thank the many people with whom I have discussed the issues, the Jensen article, and my projected criticisms. While it is impossible to list each one, I would like to mention Research Associates John Dill and Peggy Newton who have been particularly helpful and hardworking. I should also like to thank Professors Anne Anastasi, Irving Gottesman, Thomas Pettigrew, John Rainer, and Reuven Feuerstein (of Jerusalem) for supplying specific data or clarifying particular points. I have had continuing helpful dialogue with many University and Institute colleagues, particularly with Professors Martin Whiteman, Florence Schumer, Walter Neff, Cynthia Deutsch, Mrs. Alexandria Church and Mrs. Shirley Fischler. Mrs. Caroline Weichlein of SPSSI has been especially helpful. Thanks for efficiently handling a long manuscript in a short time are due Templeton Jones and Jayne Matthews and others at the Institute. Preparation of the manuscript and utilization of time of assistants were aided by small grants from the Anti-Defamation League and from SPSSI. It must be stressed, however, that responsibility for the statements contained in this article is mine alone, and that the people named above might not necessarily even have read the manuscript.

Harvard Educational Review Vol. 39 No. 3 Summer 1969

in a torrent of commentary, rebuttals, and related articles. Because of the publication lag in professional journals, most of this response appeared in the popular press and in general media such as *The Saturday Review, The New Republic,* and *U.S. News and World Report.* With the possible exception of the articles which appeared in the Spring issue of *HER*—and which were written before the publication and attendant publicity of the Jensen article—there has been no discussion to date which puts the Jensen argument and the commentaries it has provoked into a full psychological and social science perspective.

The conclusion is inescapable that the central theme of the Jensen piece is a wholly anti-democratic eugenic position, and this is dealt with at length later on in this discussion in an assessment of Jensen's concept of two ability groupings (his Level I and Level II). Thus, this relatively brief article will deal broadly with some of the specific issues raised, the arguments advanced, and the implications drawn, rather than focusing on a point-by-point discussion and refutation of errors.

I should like to make it clear at the outset, however, that in Jensen's article I found many erroneous statements, misinterpretations, and misunderstandings of the nature of intelligence, intelligence tests, genetic determination of traits, education in general, and compensatory education in particular. A colleague reports coming across 17 such errors in a casual perusal. For example, on pages 86-87, a 68% gets transposed into an 86%; on page 87 a study (Dustman & Beck, 1965) is reported with a .80 heritability factor in EEG patterns, but what is omitted is the fact that the subjects were identical twins. Perhaps so large a number of errors would not be remarkable were it not for the fact that Jensen's previous work has contained so few, and, more malignant, all the errors referred to are in the same direction: maximizing differences between blacks and whites and maximizing the possibility that such differences are attributable to hereditary factors.

In addition, in many of his citations of the literature, Jensen gives only part of the data or interpretation, or leaves out a piece of information which is crucial to his own interpretation. He also tends to use selective and sometimes inappropriate sources.

The Nelson and Dean study (1959) cited by Jensen on page 87, for example, relies on an analysis of brain wave patterns on newborn infants. Since the science of electroencephalography has yet to develop a stable picture of normative patterns in infancy, such findings are, at best, highly tentative. Interpretations from the Nelson and Dean study are further qualified by the fact that the statement about African newborns is based on only eight subjects and that the authors themselves caution the reader to treat the results with "reserve" (781).

Happenings on the Way Back to the Forum

MARTIN DEUTSCH

An example of Jensen's use of sources unsuited to a scholarly publication is his citation of a 1968 study from *Medical World News,* which he uses to suggest a link between an aspect of brain waves and IQ. The *Medical World News* consists of popularized abstracts, not scientific papers. Another example is his use of the *U.S. News and World Report* as the source for Armed Forces Qualification Test data. The figures presented in a popular publication may be correct, but an article in such a magazine cannot possibly include the subanalyses and collateral data which determine the meaning of the central test scores, to say nothing of the environmental and historical conditions which initially differentiated the populations.

Before continuing with this critique, I should like to add a personal note. I have known Arthur Jensen and respected his work for many years. He was a co-editor with Irwin Katz and me of a SPSSI-sponsored volume (*Social class, race, and psychological development*) in which the orientation was diametrically opposed to his currently stated position. His own chapter in that book, "Social class and verbal learning," is a model of clear and careful exposition of his own and others' work in this complex field (and, incidentally, is quite divergent in orientation and conclusion from his *HER* article). I am publishing this critique because I believe the impact of Jensen's article was destructive; that it has had negative implications for the struggle against racism and for improvement of the educational system. The conclusions he draws are, I believe, unwarranted by the existing data, and reflect a consistent bias toward a racist hypothesis.

I have a special responsibility to contribute to the correction of the conclusions and their foundations for two major reasons: (1) my current position as President of the Society for the Psychological Study of Social Issues, an organization dedicated to careful evaluation and interpretation of socially relevant data, in the interests of the best utilization of social science information and understanding for the betterment of man and his society;[1] and (2) my own heavy involvement in scientific and professional work related to the issues Jensen raises—the role of environment in behavior and intelligence, stimulation of intellectual development, and general compensatory and intervention efforts—which has consistently led me to quite opposite conclusions from Jensen's about the processes involved in the acquisition of knowledge, the functional dynamics of intelligence, and the severe limitations of a psychometric approach to the description of intellectual performance in human populations.

[1] The opinions expressed here are mine, however, and do not necessarily reflect the views of the organization or its members.

At the same time that I deplore the nature, conclusions, and effects of Jensen's article, I support the right of free inquiry into *all* issues, popular or unpopular. Arthur Jensen has been a consistently careful and dedicated behavioral scientist who has made substantial contributions to the study of children's learning, and especially to verbal learning. In fact, it is hard for one who has followed and read his previous work to believe that he wrote the *HER* article. One must deplore and reject the many *ad hominem* criticisms to which he has been subjected. There are enough issues raised and arguments presented in his article to provide concrete bases for disagreement and the presentation of an alternative point of view. In the critique which follows I have attempted always to remain on an *ad verbum* level.

An Invitation to Misunderstanding

Jensen's article takes the basic position that intelligence test differences between groups—most particularly between black and white groups—are reflections of differences in genetic endowment. Since the average scores of blacks are rather consistently below the average white scores, his conclusion is that these presumed genetic differences operate to make blacks inherently less competent. Contrary to the impression given by the mass media, Jensen offers no new data to support this position, but only a reorganization of existing old data. (It is important to remember that the data are mostly psychometric and not experimental or genetic.) He does add some of his own work on associative versus conceptual learning, on the basis of which he concludes that black children are more capable of concrete learning than of learning by abstraction. The policy implications he derives from this conclusion involve different curricula for black children and different expectations of their eventual intellectual level. Jensen includes numerous caveats with respect to not assuming a certain level on the part of any given *individual* on the basis of the known *group* differences, but he does not include any suggestions as to how one can identify a potential conceptual thinker in early childhood other than by his skin color.

In our present rather explosive social climate in the United States, it is not surprising that the publication of this argument and these views by a respected professor of education with extensive experimental productivity has been met by a storm of emotions and rhetoric.

In general, the published popular commentaries on the article have accepted most or all of Jensen's assertions regarding intelligence, many of his statements

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

about the measurement of intelligence, most of his genetic discussion, and with only a few demurrers, his verdict on compensatory education. Thus, James Cass in *The Saturday Review* states, "An impressive study of the nature of intelligence, its sources, and its implications for school and society was published last month in the Winter issue of the *Harvard Educational Review*. . . ." While Cass goes on to indicate that, "Dr. Jensen has presented his case, but the jury of his professional peers is still out," nevertheless the impression is created that the article is a fair and lucid discussion of the issues. In fact, however, the article falls into serious contradictions in a number of places, and completely lacks a sophisticated understanding of the magnificent complexity of environment-organism interactions.

An important consequence of Jensen's article has been to focus attention on the role of social scientists in interpreting behavior. The article has also highlighted the implications of such interpretations for formulating social and educational policy. The responsibility thus assumed by social scientists is a grave one. The Society for the Psychological Study of Social Issues (SPSSI) released a statement about the issues and arguments advanced by Jensen which dealt in part with this matter.

The statement concluded with the assertion that the Council of the Society

> ... reaffirms its long-held position of support for open inquiry on all aspects of human behavior. We are concerned with establishing high standards of scientific inquiry and of scientific responsibility. Included in these standards must be careful interpretation of research findings, with rigorous attention to alternative explanations. In no area of science are these principles more important than in the study of human behavior, where a variety of social factors may have large and far-reaching effects. When research has bearing on social issues and public policy, the scientist must examine the competing explanations for his findings and must exercise the greatest care in his interpretation. Only in this way can he minimize the possibility that others will overgeneralize or misunderstand the social implications of his work.

One major aim of the present article is to evaluate Jensen's report in the context of the foregoing consideration: it is my belief that, among other major weaknesses, Jensen's article did not demonstrate sufficient cognizance of these principles, and that the implications he draws, and most particularly the practical suggestions he makes, go far beyond what is warranted by the data he presents—or by our present state of knowledge in these areas.

One of the most forthright statements in this area was made by a geneticist,

Dobzhansky. In the context of affirming the rights of scientists to free inquiry and free expression of views he stated:

> The opinions uttered by scientists are, however, prone to be utilized by politicians and propagandists for purposes of their own. Is a scientist accountable for misuses of his discoveries and utterances? He ought to be articulate enough at least to disown such misuses. (1968, p. 129)

In exploring the implications of research on racial differences, Chicago historian Mark Haller notes: "We should not be so naive as to believe that findings on racial differences will have no policy implications in the major domestic issue that now faces the United States [1968, p. 224 f.]."

Losing Sight of the Individual

While Jensen repeatedly indicates that decisions about individuals should not be based on conclusions drawn from group data, the educational implications of his thesis prevent the drawing of this distinction between groups and individuals. Having developed the notion that white and black children tend to differ in their learning abilities according to particular parameters, which he designates Level I (associative, or rote learning) and Level II (conceptual), he then advocates differential teaching for the two groups. Despite his statement that, "The reality of individual differences ... need not mean educational rewards for some children and frustration and defeat for others [p. 117]," it is hard to understand how differential teaching for children, grouped early in life on the basis of type of learning ability (Level I and Level II), can lead to flexibility of cognitive types or teaching procedures. Further, it would be doubtful that a child taught consistently by associative, rote techniques would be able to shift to a situation in which instruction was carried out by conceptual methods. This is a critical point for occupational and status advancement, inasmuch as the greater rewards in an advanced technological society go with the more conceptual work. To assume that rote-learning and conceptual-learning groups could be maintained without status attributions and implications as simply a part of "diversity rather than uniformity of approaches and aims [p. 117]," would imply a highly naive view of the social milieu.

Jensen seems to equate his Level I and Level II with different learning styles, or patterns of ability, almost as cognitive styles, even while he designates them hierarchically as I and II, with the latter subsuming advanced cognitive and con-

ceptual abilities. If they were only styles, then there would be little reason to assume that even the child who learned scholastic skills by Level I methods (Jensen believes that Level I children can learn all the basic scholastic skills) could not perhaps go on to use those skills conceptually to solve other and more complex problems. If Jensen is referring to cognitive style, then it is likely that there are more than two styles; consequently it would be necessary to develop many different educational strategies to meet the needs of individual children. It would seem that the style notion is introduced only to make more palatable the lengthy prior argument of a dichotomy in learning ability (higher and lower) which demands differential educational organization amounting to segregation on the basis of presumed genetic inheritance.

Jensen completely neglects the failure of the school system or the larger society to achieve mass success in teaching even the basic scholastic skills. His lengthy critique assumes that potential or actual inputs are received by the child and that they get through the complex maze of environmental disorientation, scholastic chaos, and inadequately prepared teachers to a receptive organism. In essence, he fails to acknowledge the role of the school environment, the complexities of the educational system, and of the interpersonal dysfunctioning that typically characterizes the relationship of the school administration to the teaching staff, the teaching staff to the children, and inversely, of the children to their teachers. At an early age, children, often with considerable intuition and great intelligence, learn not to cope with the school situation, not to attend, not to take it seriously. In other words they find it intellectually non-stimulating, non-motivating, and in circumstances where children and teachers come from different social class and caste backgrounds, children are likely to find the interaction as well as the instruction threatening to their ego structures and personal identities. This is true for normative circumstances; it is most objective and descriptive of ghetto situations.

As I pointed out several years ago:

> ... middle-class people who work and teach across social-class lines often are unable to be aware of the negative aspects of the middle-class background because of its apparent superiority over the less advantageous background provided by lower-class life. We really have no external criterion for evaluating the characteristics of a milieu in terms of how well it is designed to foster development; as a result, we might actually be measuring one area of social failure with the yardstick of social catastrophe. (Deutsch, 1967, pp. 40-41)

With the paucity of funds available for so-called compensatory education, we have never really had a national compensatory effort. We simply must face the

grim truth that while we have had social destruction and urban decay, our overall thrust as an organized society has placed our major resources in the arena of war rather than in the area of improving general social organization, teacher training, equipment, school structures, and meaningful administrative and community participation. It would be more possible to supply both educational systems and children with relevant reading materials and the new technological aids, as well as with better trained teachers, and pre- and paraprofessionals it our priorities were reoriented toward social evolution. All these constitute requirements if any real effort is to be made toward the enhancement of the intellectual growth of the child. Until such an effort is made, it is simply not possible to arrive at a verdict as to the efficacy of education, to say nothing of the efficacy of compensatory programs.

Successful Environmental Intervention

As part of his general discounting of the effects of compensatory education programs for disadvantaged children, Jensen attributes the positive results obtained by our Institute for Developmental Studies demonstration program to the selection of samples not representative of a truly disadvantaged population (p. 98f). He points out, correctly, that the experimental sample is composed of children whose parents volunteered them for the program. (Indeed, can one ever operate a program for four-year-olds living at home unless their parents agree to it?) He hypothesizes that, "Parents who seek out a nursery school or volunteer their children for an experimental preschool are more apt to have provided their children with a somewhat better environment than would be typical for a randomly selected group of disadvantaged children [p. 98]." He fortifies his assumption that it is the self-selection that makes the difference in the Institute program by noting some data on the program indicating that the experimental (E) group and the self-selection control (Css) group did not differ significantly on Stanford-Binet IQ at the end of the kindergarten year.

Several points, both general and specific, need to be made about this reasoning and the data Jensen used. First, because we were concerned that self-selection would result in an atypical sample, we formed the Css group. This was done by selecting a larger group of four-year-olds than could be included in the experimental program and then randomly assigning them to experimental and control groups. However, at the time these groups became eligible for kindergarten, a second control group—Ck—was selected, consisting of children coming to school

for the first time in the kindergarten year. On the necessarily gross social class categorization measures, the groups do not differ from each other.

Lest the term "self-selection" give rise to a misunderstanding about how the groups were constituted, it should be made clear that the situation was not one in which there was simply formal announcement that applications would be considered. Rather, doors were knocked on in central ghetto areas, school, church, and social groups were contacted; posters were placed in various community shops and facilities. While intangibles in home atmosphere necessarily could not be measured, care was taken to insure inclusion only of children whose families could be classified at the bottom of the socioeconomic class (SES) ladder, as measured on the IDS 10-point SES scale and subsequently trichotomized Index (see *The Disadvantaged Child,* especially chapters 15 and 17).

The data Jensen referred to arise only from the second group, or wave, which was in the Institute's experimental program. It is interesting to note that Jensen failed to use the first wave data, which demonstrated significant differentiation in favor of the experimental group. Later analyses indicate that results on that second wave showed significant differences in favor of the experimental group.

TABLE 1

Anal,·is of IQ Data for Four Waves of Experimental, Self-Selected Control, and Kindergarten Control Groups in the IDS Experimental Compensatory Education Program

	Time of Test								
	Pre-Prekindergarten			Post-Prekindergarten			Post Kindergarten		
Group		Stanford-Binet	Peabody Picture Vocab.		Stanford-Binet	Peabody Picture Vocab.		Stanford-Binet	Peabody Picture Vocab.
	N[a]			N			N		
E	274	92.31	68.14	260	99.17	81.93	184	100.49	88.14
Css	129	91.37	65.50	142[b]	92.08	72.38	98	93.17	80.82
Ck	—	—	—	180	90.95	72.39	177	92.50	81.22

[a] N's differ because of attrition. A special analysis indicated that there was no relationship between attrition and IQ score, indicating that attrition was not selective in terms of IQ score.

[b] Increase in N's is because only half of the first wave group was pretested to control for possible effect of pretest on posttest. (There was no effect.)

(There was an error in the report of these data in a non-scientific, popular publication by Powledge, cited by Jensen.) We have recently completed an analysis in which results of four waves of experimental and control groups were examined. These are reported in Table 1.

These data refer to several waves or groups and reflect a much more substantial *N*. They indicate that though the E and Css groups do not differ significantly at the time of pretest they do differ significantly at the end of the nursery year and at the end of the kindergarten year. Thus, the program has an effect independent of the self-selection variable. In addition, the Css and Ck groups do not differ significantly, either before or after the kindergarten year. Since the Ck group is randomly selected from entering kindergarteners in the same schools in which the experimental classes are located, it appears that the factor of self-selection for the experimental program did not produce a group of subjects atypical of the disadvantaged population in the neighborhood of the school. Therefore, Jensen's argument that the E and Css samples are not representative does not hold.

Other test results (Lorge-Thorndike, Illinois Test of Psycholinguistic Abilities) on these samples confirm the positive findings with respect to the effects of a longitudinal enrichment program. On the data from the Illinois Test of Psycholinguistic Abilities (ITPA), for example, differences between E and Css groups were greater in the third grade than at the time of earlier testing. Data analyses reveal significant differences in favor of the experimental children on the ITPA total score, and on six of the nine subtests. On the other three subtests, the experimental group scored higher, but the differences were not statistically significant (C. P. Deutsch & C. Silfen, 1969). Analysis of results from recent testing (Spring, 1969) with six subtests of the revised longer ITPA is confirming these earlier findings. It appears from the subtest pattern that intervention specifically may help to counteract initial deficiencies in the auditory and vocal modalities, thereby enhancing development of verbal association and communication. In addition, recent data from the Metropolitan Reading Test indicate that reading scores of experimental children are at, or above, national grade average at the end of third grade, and are significantly different from control group scores.

It would appear, at least from the results of the Institute's program, that Jensen has prematurely classified compensatory education as a failure. The findings briefly reported here clearly demonstrate that continuous and carefully planned intervention procedures can have a substantially positive influence on the performance of disadvantaged children and avoid the cumulative failure all too fre-

quently found. The Institute program and data have been used for illustration, as I am most familiar with them. However, there are other effective compensatory programs which have been reported elsewhere. Some of these are discussed explicitly in the recent report by the American Institutes for Research (1969).

Long-Term Programs Are Needed

No doubt one factor that led Jensen to an erroneous conclusion is that focused compensatory education for disadvantaged children is a quite recent development, and early reports of results necessarily came from shorter and usually more hastily conceived and poorly handled programs, on both federal and local levels. It should come as no surprise that children born into poverty and all it implies cannot be rescued in an isolated summer or even by a year of a nonstimulating school program. When one considers the magnitude of the problem, especially when deprivation is confounded by the effects of discrimination suffered by many minority groups, it is hardly surprising that programs such as summer Head Start generally failed to have any lasting influence on the lives of disadvantaged children. However, the above data from the Institute's program, and those from other long-term efforts, indicate that long-range enrichment with specially trained teachers, careful planning and supervision, and adequate funding can produce positive effects on IQ scores, on specific language skills, and on school achievement measures. Even though it is not yet possible to tell what the longer-term effects will be (e.g., on high school performance and on adult occupational status), the current results are encouraging indeed, and are more than sufficient to reject the blanket conclusion that compensatory education has failed (if one assumes that it has ever really been attempted). On the contrary, what is strongly indicated is the establishment of more long-range, continuing programs for children from the slums. Careful evaluation of results of varying programs will yield information as to the best operative procedures and should lead to more efficient and more effective compensatory education, even in the context of an increasingly dysfunctional and irrelevant school system.

Unfortunately, Jensen apparently has a somewhat idealized view of the school system. He says, "The interesting fact is that, despite all the criticisms that can easily be leveled at the educational system, the traditional forms of instruction have actually worked quite well for the majority of children [p. 7]." This makes Jensen one of the few professional observers who would defend the current school system and the opportunities it offers for both specific skill development

and more broadly defined intellectual growth. He takes the position that the curriculum is organized in a way that demands and fosters abstraction, problem solving, and concept formation. In actuality, schools are oriented far more to associative or rote learning, as can be seen in workbooks and sample lesson plans, as well as in the over-structured, non-creative, non-responsive classes that typify most of America's schools. For the black ghetto child, Kozol's *Death at an Early Age* is a much more accurate rendering of the objective school experience. In Lewinian terms, one might say that the black ghetto child's life space and opportunity for independent behavior are rather harshly restricted, and in actuality often reflect a behavioral rendition of the desolate landscape of the moon.

Jensen doubts that IQ can be much affected by environmental means, other than environmental effects *in utero*. Such a position appears unwarranted, in view of lines of evidence from sources other than the previously discussed reports of positive effects of some compensatory education programs.

Some of the most interesting work on the modifiability of intellectual abilities comes from studies of children in Israeli kibbutzim. Smilansky (1964) reported some of the early data which were also discussed in a 1964 conference. (For a report of the proceedings of the conference, see Hess, Davis & Bloom, 1965.) Particularly dramatic are data showing changes in the IQs of Oriental children after four or more intensive years in the kibbutz nursery. Bloom (1969) refers to findings that under these conditions the IQ levels of Oriental children rose from a mean of 85 to a mean of 115. The direction of change, although not its magnitude, is consistent with the early reports of Klineberg (1935) and the later study by Lee (1951) which demonstrated an increase in the mean IQ of southern Negroes who migrated to the North. Both lines of data indicate the role of environment in modifying IQs, with the differential magnitude of change undoubtedly attributable to the very different levels of fostering conditions in kibbutzim and in northern American cities.

The work of Feuerstein (1968) with retarded children in Israel casts further doubt on Jensen's view that environment has little effect after the child is born. Feuerstein has shown that, with adequate stimulation, many children who initially show a low level of functioning (comparable to Jensen's Level I) can reach a much higher level of functioning (similar to Jensen's Level II). Considering Jensen's statement that he had found "... no studies that demonstrated gains in relatively non-cultural or non-verbal tests like Cattell's Culture Fair Tests and Raven's Progressive Matrices [p. 101]," it is pertinent to note that one of Feuerstein's measures on which gains were noted was the Raven. The magnitude of

gains reported by Feuerstein and others is so substantial that question must be raised as to the even elementary adequacy of our own current intervention models. In this area it would appear that Jensen has inverted his periscope and is looking at the wrong answers, as well as at the wrong questions.

Extrapolations, Contradictions, and Misinterpretations

Jensen relies very heavily (especially on pp. 84-87) upon the Coleman report (1966) to indicate that situational-environmental factors are not of essential importance to school achievement. He refers to two studies—the only two he characterizes as "methodologically adequate"—of father absence. Both studies, he says, indicate that the father factor does not contribute independently to variance in intelligence or scholastic achievement. It seems somewhat incredible that one of the two studies he finds "methodologically adequate" is the Coleman report, inasmuch as this is one of the most massively criticized reports issued in recent years, with the bulk of the criticism centered on its methodological inadequacies (e.g., Jencks, 1968).

One problem in the Coleman report comes from the fact that there was a substantial differential response rate to the questionnaires on which it is based. In numerous categories there was a return of less than 50%. In addition, the data suffer from a great unevenness, as they were gathered by means of questionnaires filled out by school administrators, teachers, and others of varying levels of involvement, understanding, and sophistication. Most of the questionnaires were sent and returned by mail, which further added to the differential return and validity. In a limited number of instances, the data were gathered by untrained interviewers working with a questionnaire that was unfamiliar to them and which demanded that they ask probing questions as to reading material in the home, cultural amenities, preschool education, parents' education, child's self-concept, and so on. It is not my purpose here to discredit the Coleman report, but only to delineate the controversy which has surrounded it, and to which Jensen gives us no clue. It is almost impossible to make valid generalizations from the Coleman report which was hastily conducted and included numerous methodological difficulties. Any social scientist who chooses to use these data in support of his position must at least acknowledge the methodological problems and the limited scope of the data. In his use of the Coleman data, however, Jensen demonstrates an absence of understanding these limitations.

Jensen makes another fundamental error in equating social class across caste lines, as if black middle-class experience were identical with white middle-class experience. Actually, it may take as much as two or three generations of real middle-class status before a black man will be able to have the kind of socializing advantages that most white people in our society enjoy today. This means that it is impossible to verify or validate most of the studies that have been done in this area except to look at them as interesting operations in terms of the first encounters of social scientists with the complexity of the human experience and human organization.

In an important review of literature, comparing the performance of Negroes and whites, Dreger and Miller (1960) state that it is not enough to equate ethnic groups in terms of social class and economic variables; that there is a caste as well as a class difference; that Negroes, with earnings equal to or better than whites, will still typically be prevented from living the same kind of life. This conclusion is stated in the context of Dreger and Miller's explicit statement that they take no sides in the so-called traditional heredity-environment controversy.

Citing both Coleman (1966) and Kuttner (1967), Jensen claims that American Indians are considerably more disadvantaged than black Americans or other minority groups (p. 85). The Kuttner data did not come from the Coleman study, and therefore may or may not be using comparable samples with respect to income and unemployment statistics. What these data basically indicate is a greater degree of *structured* environmental deprivation within the Indian community than within the ghetto. Jensen says, "... the American Indian ability and achievement test scores average about half a standard deviation higher than the scores of Negroes [p. 85]" and that "... differences were in favor of the Indian children on each of the four tests used by Coleman: non-verbal intelligence, verbal intelligence, reading comprehension, and math achievement [p. 85f.]." Then Jensen submits, "If the environmental factors assessed by Coleman are the major determinants of Negro-white differences that many social scientists have claimed they are, it is hard to see why such factors should act in reverse fashion in determining differences between Negroes and Indians ... [p. 86]." Such a question simply ignores the problem of measuring the salient or operative variables within any disadvantageous situation and relating them to criterion measures. What is implied by the question is that all disadvantage is essentially the same, and exists only in differing quantities. Actually, of course, it is impossible to avoid recognizing that there are qualitative differences between environments, and that these are probably highly relevant to any discussion of environment-be-

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

havior relationships. For example, in superficially comparing Indians and Negroes, Jensen completely ignores the special conditions of American Indians: their history, their current social organization, and their schooling.[8]

Perhaps more important than Jensen's oversimplification of the Coleman data, or his ignoring evidence of success through compensatory education, is his attempt to generalize from the classic heritability studies. They are, after all, studies of Caucasian children, especially separated twins or siblings, whose *environmental* variation is not thought by scholars to be representative of the general population. To say it bluntly, Jensen (and the rest of us) have no idea what the proper estimates of V_H, V_E and so forth are for black people, and we have only very tentative guesses as to what they are for Londoners and northern mid-west Americans. The estimates of heritability, upon which Jensen's entire argument depends, are only accurate if each possible genotypic child is placed randomly in each conceivable environment. To approximate such a study, researchers must at least include black children and a representative range of environments.

If we take into consideration a number of factors discussed on different pages of the article, we find that Jensen destroys his own main argument. He explicitly states that the median IQ difference between Negro and white samples is 15 to 20 points. If we add the 8 or 10 points attributable to the test situation, the few points which Jensen concedes can be gained in compensatory education, and the additional 5 points which he is willing to attribute to poor environments, we find that all statistically significant differences have been obliterated. Jensen thereby leaves himself with no argument.

It is this kind of conflicting and contradictory reporting that makes it very difficult to take the Jensen article seriously in either scientific or logical terms. It is tragic, therefore, that its conclusions have been so widely disseminated by the mass media.

Another example of Jensen's misinterpretation of his own data is to be found on page 83. In his analysis of his own table on the prevalence of retarded children by race and SES, Jensen says, "If environmental factors were mainly responsible for producing such differences, one should expect a lesser Negro-white discrepancy at the upper SES levels." In examining the table, if we look at percentage differences between Negroes and whites at each SES level, we find a difference of 2.6% at the highest SES level and 35.1% at the lowest SES level. This

[8] For an extensive discussion of test variables with respect to different groups of subjects, see Deutsch, M., Fishman, J., Kogan, L., North, R., and Whiteman, M., Guidelines for testing minority group children. *Journal of Social Issues*, 1964, 20 (2), 129-45.

analysis, based on Jensen's own data, supports the environmental hypothesis. However, he goes on to discuss the issue as though the table demonstrated the reverse: he is consistent with his bias but not with the data.

Hefner (1969) criticizes the logic of Jensen's statement on page 83: "Since in no category of socioeconomic status (SES) are a majority of children found to be retarded in the technical sense of having an IQ below 75, it would be hard to claim that the degree of environmental deprivation typically associated with lower-class status could be responsible for this degree of mental retardation [p. 4]." Hefner suggests that the statement would be equally logical if other phrases, such as "found to be undernourished," or "found to have only one leg," were substituted for that on retardation. Even apart from this all-purpose statement, however, is the fact that differential prevalence of IQs below 75 and the probability that very low IQs are associated with neuro-biological deficits (by no means necessarily genetically determined) may have nothing at all to do with observed test score differences in the IQ range above 75.[8]

On page 62f., Jensen discusses the Wheeler (1942) data and appears to say that a decline in IQ was observed in a longitudinal study. Again, I quote Hefner,

> ... but in fact there is only a 1930 and a 1940 cross section. Thus, when he says that the 'decline in IQ from age 6 to 16 was about the same in 1940 (from 103 to 80) as in 1930 (from 95 to 74).' What he really means is that separate samples from the group which averaged 95 in 1930 at age 6, averaged 80 in 1940 at age 16—after some years of state and Federal intervention to improve the environment of the area. There is no group that 'declined' from 103 to 80, or from 95 to 74. (p. 4)

In another seeming contradiction (p. 100), Jensen states that he would put little confidence in a single test score, and especially if it were a child's first test score; he adds his limited confidence in the result if the child is from a poor background and of a different race from the examiner. On page 108, Jensen points out that educators should de-emphasize IQ scores as a means of assessing gains and use mainly direct tests of the skills the instructional program was intended to teach. Despite this cautious view of IQ tests, however, Jensen gives us 100 pages of interpretation of IQ test results in terms of race, genetic determination, teaching methods, and general environmental influences.

As evidence for his conclusion that middle-class white children do better than lower-class black children on conceptual (Level II) tasks, Jensen relies heavily

[8] Actually, 75 is an unusual cut-off point in the mental retardation literature. Typically, 68 or 70 is used. The proportion of cases between 68 and 75 is not given in Jensen's report, but is usually substantial.

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

on a study by Glasman (1968). Jensen describes the results as indicating that middle-class children do better on recall of objects, which can be clustered into meaningful categories, than do lower-class children. He relates this to a previous finding of his own that SES differences are not apparent in free recall of unrelated objects. However, he does not indicate if the two samples were comparable; he does not even give the age (s) of his sample. The critical importance of age is clear from the report of the Glasman study, which found no SES differences between recall on categorized and uncategorized lists in kindergarteners, while differences were present for fourth and fifth grade children. Thus, while Jensen reports the Glasman study as a kind of extension of his own earlier work, and as support for his Level I-Level II differentiation, the age-related differences would have to be compared for the two studies before any conclusions could be meaningful.

In trying to explain his own observed finding that Level I tasks correlate with IQ among middle-class children but not among lower-class children, Jensen postulates a scatter diagram of correlations within class groups. He says, "Since large representative samples of the entire school population have not been studied so far, the exact form of the correlation scatter diagram has not yet been well established, but the schematic portrayal of Figure 18 is what could be most reasonably hypothesized on the basis of several lines of evidence now available [p. 113]." Since he does not specify the "several lines of evidence now available," what he has apparently done is to construct two diagrams that would reflect his findings without destroying his conclusions. There is, thus, no apparently valid relationship between the scattergrams and reported data. However, interposing them between the stated finding of high Level I task/IQ correlations among middle-class children, and low correlations for lower-class children, and his later statement that "Level I ability is distributed about the same in all social class groups, while Level II ability is distributed differently in lower and middle SES groups [p. 114]" might have the effect of making the already tired reader ignore the inconsistency of the two statements.[4]

Jensen's postulation of Levels I and II—separate associative and conceptual intellective processes—cannot be seriously considered from a theoretical point of view. If one were to draw on current intellective and behavioral theories, there would be a basis for a theory of intelligence founded on a total interpenetration of cognitive and associative levels. I would postulate further a third level, which would

[4] The reader who goes back to check this should not be confused by an obvious proofing error: the captions for Figures 18 and 19 are reversed in most issues, correctly placed in the reprints.

subsume the other two and include as well the organism's own personal experiences and history: its deprivations and reward systems. These systems embody as well an internalize responsive network that creates a self-reinforcing organismic individuality, which would constitute the psycho-behavioral level of the self-fulfilling prophecy. (A fuller discussion of this construct is included in a paper to be published in the winter issue of *The Journal of Social Issues.*)

The Eugenic Tautology

In evaluating Jensen's dual cognitive typology of intellectual performance, it is necessary to read carefully his discussion on page 114. He says:

> That learning is necessary for Level II no one doubts, but certain neural structures must also be available for Level II abilities to develop, and these are conceived of as being different from the neural structures underlying Level I. The genetic factors involved in each of these types of ability are presumed to have become differentially distributed in the population as a function of social class, since Level II has been most important for scholastic performance under the traditional methods of instruction.

This is perhaps the clearest statement of the position which is fundamental to Jensen's total argument. It is quite similar to Shockley's request to the National Academy of Sciences (1966) in which he suggested that the Academy undertake a major investigation of the possible genetic determinants of racial differences in intelligence. In the Academy's most recent rejection of Shockley's proposal (in which Shockley cited Jensen's *HER* article) Dr. Frederick Seitz, the President of the National Academy of Sciences, was quoted as saying, "It is essentially impossible to do good research in this field as long as there are such great social inequities." Dr. Seitz based his position on the Academy's policy statement (NAS *News Report,* November, 1967), which holds that it is not clear, despite the tests which have been done, whether differences in intelligence between the black and white populations are genetic or environmental and that there is no scientific basis for supposing them to be either one or the other. Subsequently, in an even stronger statement clarifying the Academy's view, Seitz said, "There is a strong feeling within the Academy that social inequities make it impossible to do reasonable scientific research in this area.... In addition, the conduct of such research at the present would tend to heighten current social tensions to a very destructive degree [1969, p. 652]."

Jensen's later discussion of his dual cognitive typology of intellectual perfor-

mance (Levels I and II) is not consistent with his earlier apparent characterization of levels of abilities as "styles." In the statement on page 114, Jensen assumes that there are different neural structures characterizing Levels I and II and that there are important genetic factors involved in determining these structures. This is a restatement of the old Galtonian eugenic point of view, which essentially hypothesizes high positive correlations among social class, intelligence, and neural factors. The social implications of this are enormous, obvious, and totally anti-democratic, and would tend to create a permanent caste society in which those of lower caste (mostly black) would be forever doomed by their hypothesized neural structures to remain in an inferior position, with all that it implies for future occupational attainment and the antecedent educational opportunities.

The impossibility of linking genetic factors with racial factors, social factors, and intelligence is described by Fried (1968):

> Absolutely no study yet done on a so-called racial sample of human population adequately links intelligence, potential ability, educability or even achievement to a specifiable set of genetic coordinates associated with any aggregate larger than a family line or perhaps lineage. (p. 124)

Scott (1968) further points out:

> ... the range of human adaptation is so great that it is doubtful whether population differences on any behavioral test of complex performance ever can be assigned to any definite genetic basis. (p. 65)

I think it is of primary importance in this discussion that we recognize that there is no built-in correlation between IQ test measurements and the nature of intelligence. They are quite different, and, unfortunately, Jensen's article continually translates one into the other. In spite of disclaimers, he constantly uses the terms interchangeably, and the general reader comes out with the impression that an IQ score and intelligence are synonymous.

What Chein pointed out in 1945 is still true:

> No psychologist has ever observed intelligence; many have observed intelligent behavior. This observation should be the starting point of any theory of intelligence, but such has, unfortunately, not generally been the case. (p. 111)

With respect to intelligence testing, it would seem that we are deluding ourselves if we believe that such tests truly indicate something about capacity or

about general learning ability, or that they even reflect a child's current cognitive skills, to say nothing of predicting his potential skills, especially if facilitating stimuli are given, such as Blank (1968), Feuerstein (1968), Caldwell and Richmond (1968), Deutsch (1967), and others have demonstrated.

IQ Is Not a Measure of Capacity

Standard intelligence tests measure essentially what children have learned, not how well they might learn something new. Intelligence tests have been constructed within a certain kind of society and a certain kind of cultural milieu, basically white middle-class America. During a period of dynamic social change, tests have remained static and have become increasingly irrelevant for understanding the nature and evolution of an organism's intellectual behavior.

Chein, later in his important article on the nature of intelligence, states:

> Psychologists who are keenly aware of the fallacy of reification with respect to other concepts and even those who have in their discussions of intelligence, often enough, verbalized the danger of hypostatizing entities where none exist have, nonetheless, tended to ascend the ladder of abstraction so rapidly that they have often left the fundamental observation far behind. (1945, p. 111)

Arthur Jensen has committed this error in his rapid ascent from test results to heritability formulas for "intelligence."

Early in the paper, Jensen introduces the concept of *g*, which designates the theory of intelligence proposed by Spearman (1923). It refers to the notion that all intellectual activity partakes of a common, *general* (*g*) factor. Jensen's subsequent discussion of intelligence and intelligence tests, including his definition of Levels I and II, is based on the *g* theory: he defines tests in terms of how much *g* loading they have, and describes his Level II intellectual functions as *g*.

However, *g* represents only *one* theory of intelligence, among many others. It is by no means a universally accepted concept among psychologists and others who work in this area. Yet from Jensen's paper, the general reader would never know that there are competing theories, several of which are more widely accepted and based on more recent information and data than Spearman's.

Spearman's theory stemmed from the early development of factor analysis. Thurstone subsequently developed the technique of multiple factor analysis, and from his studies derived a multi-factorial theory of intelligence (1938). Thurstone's theory regarded intelligence as being composed of a number of different factors, which did not have to bear any specific relationship to each other. While

Thurstone allowed for the potential existence of *g* as a structural sub-strate, it was not an intrinsic part of his theory.

A prime example of the later development of theories of intelligence based on factor analysis is Guilford's theory of the structure of intelligence (1959, 1967). This theory is the culmination of many years of work in the area by Guilford and his associates. The picture of intelligence generated is a multi-factorial, multi-faceted one that Guilford believes is reflective of the actual complexity of human beings. He says:

> There are many individuals who long for the good old days of simplicity, when we got along with one unanalyzed intelligence. Simplicity certainly has its appeal. But human nature is exceedingly complex, and we may as well face that fact.... Humanity's peaceful pursuit of happiness depends upon our control of nature and of our own behavior; and this, in turn, depends upon understanding ourselves, including our intellectual resources. (1959, p. 479)

In addition to these factorial theories of intelligence, there are various theories which derive from different lines of development. For example, Piaget's theory of intelligence derives from a developmental analysis of children's thinking (1952). One hallmark of the theory is the notion that intellectual development is intimately interwoven with the child's experiences: through the dual processes of assimilation and accommodation, the child comes to know his world, to incorporate this knowledge, and to modify his understanding in terms of new experiences and interactions. Piaget's theory is a "stage" theory, in the sense that levels of development are considered to be achieved in a fixed order, with each level building on the previous one. Whereas Jensen's notions of level are categorical and static, Piaget's reflect the idea of process.

None of these theories of intelligence has been "proven"; incontrovertible data have not been gathered to confirm any of them. However, each of the theories mentioned is as valid and prominent as *g*. Jensen's entire argument appears to be inextricably linked with the concept of *g*. Questioning *g* throws Jensen's whole line of reasoning into doubt.

Chein (1945) takes an altogether different approach in describing intelligence. He states:

> Intelligence is an attribute of behavior, not an attribute of a person. Even though we may observe some constancy in how intelligently a person acts in different situations, we may, on this basis, speak of the person's characteristic behaviors and not of a genuine attribute of the person. (p. 119)

The Attribution of Environmental Effects to Heredity

Jensen's failure to discuss other theories of intelligence and the lack of any explication of his reasons for preferring the *g* theory is consistent with his unexplained selectivity of studies, theories, and literature throughout the article. Thus, in his brief discussion of the potential effects of pre- and paranatal variables on later development, Jensen refers to the studies by Stott (1960, 1966), but ignores the massive work in the area by, e.g., Pasamanick and Knobloch and their associates (1967, 1969). The fact that Stott allows for a genetic hypothesis, while the other investigators interpret their findings in social-environmental terms, undoubtedly is a factor in Jensen's preference. At the same time, the body of work of the others is so substantial that it can hardly be ignored in any discussion of this area.

Briefly, the Pasamanick-Knobloch group found a relationship between the socioeconomic level of the mother and the incidence of pregnancy and paranatal difficulties, including prematurity (specifically defined in terms of birth weight). In turn, pregnancy complications and birth difficulties are associated with a higher incidence of neonatal mortality, morbidity and brain damage, and subsequent learning and behavioral disorders. Montagu (1967) points out that maternal nutrition, especially vitamin and protein intake, is one of the variables heavily implicated in neonatal birth weight (prematurity) and condition, as well as in the other paranatal disorders mentioned. In his discussion, Montagu indicates that even the nutrition of a child's grandmother can affect the child, since the state of the grandmother's nutrition before and during her pregnancy would have influenced the quality of the mother's ova which were later fertilized. The same factors are operative for prenatal influences on the tissues which gave rise to the sperm, which subsequently fertilized the ova. Since low SES women typically have poorer nutrition than middle-class women, social and economic variables are clearly implicated. Montagu does allow for a potential genetic factor in susceptibility to the negative effects of poor nutritional status (i.e., not all individuals or groups need be equally adversely affected by the same degree of nutritional inadequacy), but the fact remains that such (possibly genetic) differential susceptibility would be operative only in interaction with (SES-related) poor nutrition.

Until such relationships are disproven (which seems unlikely), it would appear scientifically indefensible to discard social factors as major influences on pre- and paranatal events.

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

Jensen's discounting of the importance of social factors in this area is yet another example of his insufficient appreciation of the complexity of environment itself as a variable, and of the even greater complexity of organism-environment interactions. In sections of the article in which he discards environmental hypotheses (e.g., p. 84f.) as a source of group differences, Jensen does little analysis of variables in the environment, but rather seems to regard the environment as a kind of unit. The variable he does separate out (p. 85) is a social-familial one: "father absence." Whereas in earlier writing (e.g., "Social class and verbal learning," 1968), Jensen discusses the need for "task analyses" and attempts to examine the differential verbal habits of different social classes, he does not deal with social variables on this level in his *HER* article. As a result, what emerges is a picture of some social-class related holistic environment to which is attributed only a relatively small proportion of the variance of observed group differences. There is no consideration of a process of interaction between an individual and his environment.

Mediators in a Complex Environment

An exploration of the nature and effects of such interaction is found in C. Deutsch's discussion of environment and perception (1968). Using perception as an exemplar dependent variable, she analyzes the history and transformations of the "heredity-environment" controversies and asserts that, as long as the issue was posed in such global terms, no specific data could emerge. Changing the terms to "nature" and "nurture," however, opened the way to specification of influential variables and to their hierarchization. On the basis of her analyses of both theories and data, she concludes that life conditions—including current social situations, past experiences, and cultural and socioeconomic factors—influence fundamental developmental processes. She hypothesizes that these influences operate through "mediating variables," which relate to environment on the one hand, and to behavior on the other. Referring to work in perceptual learning (e.g., Gibson & Gibson, 1955; E. Gibson, 1963; Covington, 1967) and to sensory deprivation experiments (e.g., von Senden, 1932; Hebb, 1958), she emphasizes the role of the stimulus in learning and behavior. She suggests that the conditions of life for the individual are determinants of the quantity and nature of stimuli to which he is exposed, and that, therefore, one large class of mediating variables includes the actual stimuli which impinge on the individual.

These stimulus theories are consistent with a notion of modifiability of perception, as a result of particular stimulus presentations. From this point, Deutsch draws practical implications for the organization of classroom and school materials. Considering the prime importance of both visual and auditory discrimination in early learning and in the acquisition of foundation skills, such as reading, she believes that the school learning process could be greatly enhanced by appropriate organization of stimuli, so that the child could be provided with the greatest amount of relevant practice in building his discrimination skills. Slum environments, Deutsch suggests, do not provide young children with a sufficient variety of stimuli, and most especially do not provide the kind of figure-ground, or signal-noise ratio, which is conducive to accurate and defined perception. Also, as compared with his middle-class peer, the slum child is less often told the names of the objects and noises he perceives and, consequently, he is further hampered in the development of stable discrimination skills.

Deutsch believes that the school situation can do much to remedy whatever perceptual discrimination deficiencies the child brings with him (providing, of course, he is not brain damaged or sensorily impaired). She proposes a "stimulus analysis" of classrooms and materials as a basis for formulating their appropriate organization and construction. Since perceptual processes play an important role in intelligence test performance, it is possible that remediable (and, according to these theories, environmentally conditioned) perceptual difficulties contribute substantially to observed SES differentials in IQ. Visual discrimination is an especially relevant factor in such tests as the Raven Progressive Matrices, on which Jensen places some emphasis; but he does not consider perception or perceptual development in his article. Neither does he consider the kind of operational role of the environment and its stimuli which Deutsch postulates.

In Jensen's article, heredity is similarly seen as a kind of global variable, but one which exercises a decisive influence on development. Further, this influence is seen as predetermined (from the time of conception) and as unmodifiable in its operation. The only kinds of interaction allowed for in Jensen's system are epistasis (interaction between genes) and the rather simple type of interaction exemplified by the attainment of height: the limit is set genetically, but factors such as nutrition and illness can prevent an individual from attaining his maximum stature. The implication of this position is that nature is unmodifiable, since it is considered as intrinsic to the individual. This simply means that people are locked into their individual life cages by their genetic blueprints, and environment

has no influence, except perhaps to interfere with the achievement of one's full genetic potential.[5]

My emphasis in this paper on environmental influences does not mean that I am discounting genetic factors. It is simply that they must be seen as *interacting* determinants, rather than as separate causative agents, especially when behavior is considered. For example, Hilgard and Atkinson (1967) explore the issue of the complex interaction of heredity and environment, and suggest that the methods of genetics may be applicable to behavior as well as to structure. They believe that the chromosomes and genes must be responsible for the inheritance of various components of behavior, as they are for inherited structure. However, they make it clear that they are referring to behavioral *components,* rather than to complex developed behaviors. They also point out that some genes are dominant; some recessive; and some are sex-linked, so that predictions can be made only in terms of statistical probabilities.

In considering genetic influence on traits, it is important to make the distinction between genotype and phenotype. As Gottesman (1968b) states:

> Genotype refers to the totality of factors that make up the genetic complement of an individual. Phenotype refers to the totality of physically or chemically observable characteristics of an individual that result from the interaction of his genotype with his environment. Environment must be broadly defined to include not only intrauterine and postnatal conditions but also a host of molecular factors within and between the embryonic cells (Waddington, 1957).
>
> Different genotypes may have the same phenotype, and different phenotypes may be displayed by the same genotypes. A lack of clarity is perpetuated in discussions of individual differences by a failure to specify the environmental circumstances when describing the phenotype of genes. And conversely, the attribution of an effect to an environmental manipulation may be misleading unless the genotype is specified. (p. 29)

Of course, in humans, specification of the genotype is extremely difficult, even for relatively simple traits, since each generation is so long, relatively few offspring are produced, and selective, controlled breeding is not possible. It is more

[5] As C. Deutsch points out, however, modern genetics teaches that genic operation itself is responsive to environmental variation. For example, experiments show that incubating Drosophila larvae at one temperature will produce one color of adult fruit fly, while incubating larvae from the same genetic strain at a different temperature will result in adult individuals of a different color. The environment, then, affects the biological attributes of the organism by influencing the operation of the genes.

That the temperature has not simply produced a genic mutation is shown by the fact that offspring of the two sets of larvae, incubated at the same temperature, all develop into fruit flies of the same color.

possible, though, to develop analyses and methods of specifying environmental variables. As Gottesman points out, the more similar the environments, the more variability in traits can be ascribed to genetic factors. Perhaps the best approach to determining the genetic contribution to a given trait would be to describe accurately relevant environmental variables and then attempt to subject the varying populations under study to as nearly identical environments as possible. Since this has not been done, and since Jensen must recognize the differential environmental milieus of different social class and racial groups, it seems inescapable that his main thesis of genetic structure as the major source of variance in intelligence test score differences between social class and racial groups must be rejected on that basis alone. In considering Jensen's heritability formula in the light of these facts and definitions in modern genetics, Hirsch's (1968) statement is most apt:

> Only when we consider the number of possible genotypes and the number of potential environments that may influence trait expression do we begin to realize how narrowly limited is the range of applicability for any obtained heritability measure. (p. 42)

Jensen relies heavily on kinship studies, particularly twin studies, for his estimate of the heritability of intelligence. While twin studies represent a logical and appealing approach to the heredity-environment question, they present several serious methodological problems. [For a more complete discussion of these problems, see Woodworth (1941), Essen-Moller (1963), and Vandenberg (1966).]

Fuller and Thompson (1960) point out that, "Methods of treating twin data cover a wide range of statistical procedures, some naive and others highly sophisticated [p. 109]." Unfortunately, Jensen's description of his procedure does not give the reader sufficient information to determine into which category his method falls. He may have used studies involving direct comparisons of monozygotic and dizygotic twins (a procedure to which there are many methodological objections), or he may have used twin studies employing other conditions. He states only:

> I have presented elsewhere a generalized formula for estimating heritability from any two kinship correlations where one kinship is of a higher degree than the other (Jensen, 1967a). I applied this heritability formula to all correlations for monozygotic and dizygotic (half their genes in common) twins reported in the literature and found an average heritability of .80 for intelligence test scores. (p. 51)

It would appear from this statement that he lumped together twin studies without reference to their widely differing levels of methodological adequacy.

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

Perhaps the most empirically derived twin studies in the literature on the influence of heredity on intelligence are those of identical twins reared apart. Jensen mentions three of the four existing studies. It is interesting to note that he gives most attention to Burt's (1966) study. It contains the highest estimates for heritability in this literature, and its findings are not completely substantiated by other studies.

While Jensen mentions intra-pair correlations for intelligence test scores of identical twins reared apart, he does not examine the mean intra-pair differences. In examining these differences, summarized in Table 2, we find that the average difference ranges from 6 points (Burt, 1966) to 14 points (Shields & Gottesman, 1965). Gottesman (1968) notes that at least 25% of Shield's (1962) sample of twins reared apart (N = 38) has within-pair IQ point differences of more than 16 points on one of the tests. In the studies cited in Table 2 the maximum within-pair difference ranges from 14 to 30 points. Such variation between co-twins, often significantly correlated with environmental differences, suggests the impact of environment on IQ test scores.

An example of the considerable influence of environment is seen in the frequently-cited study of identical twins reared apart by Newman, Freeman, and Holzinger (1937). In this study it was found that the IQ scores of identical twins

TABLE 2
Mean Intra-Pair Differences in IQ Test Scores

Study	*N*	*Test*	*Mean Differences in IQ Points*	*Range of Differences in IQ Points*
Newman, Freeman & Holzinger (1937)	19	Stanford-Binet	8.2	1-24
Shields (1962)	38	Combined Score[b]	9.5	0-30
Shields (1962) as reported by		Mill Hill Vocabulary	14	—
Shields & Gottesman	38	Dominoes	10	—
Juel-Nielsen (1964)	12	Wechsler-Bellevue	7.3[a]	1-14
Burt (1966)	53	Stanford-Binet (London Standardization)	6	—

[a] Computed by P. Newton
[b] From Dominoes and Mill Hill Vocabulary

who were reared apart and who were separated during the first three years of life showed a correlation of .79 with educational advantage. Bloom (1964) analyzed these data, dividing the identical twins reared apart into two groups. In one group of 11 pairs with very similar educational environments, the rank order correlation of IQ test scores was .91, in contrast with a rank order correlation of .24 for the eight pairs of twins with less similar educational environments. From this analysis, they conclude that,

> ... if the identical twins are separated but placed in very similar environments, it is likely that they will have very similar intelligence test scores, whereas if placed in very different environments, their intelligence test scores will be quite different. (1964, p. 70)

Using the Newman *et al.* ratings of educational and social differences between pairs of twins, Stone and Church (1968) classified 10 pairs of twins as having "larger differences in educational and social advantages" (DSEA), and nine pairs of twins as having "smaller DSEA." They found that seven pairs of the twins in the larger DSEA group had IQ differences of 10 or more points, while only three pairs of twins in this group had IQ differences of less than 10 points. In the group with the smaller DSEA, all pairs of twins showed IQ differences of less than 10 points. In the larger DSEA group, four pairs of twins showed differences of 15, 17, 19, and 24 IQ points.

Results also suggestive of the influence of environmental factors on IQ test scores come from Juel-Nielsen's (1964) study of 12 pairs of identical twins reared apart. Examining his results for the seven pairs of twins who had had differences in educational experience, Juel-Nielsen found significant differences ($p \leq .05$) on the following parts of the Wechsler-Bellevue: Information, Digit Span, Verbal Points, and Total Points. (Jensen did not cite this study in his article. The omission is unfortunate, as this study handles several of the methodological difficulties present in the other three studies of this design.)

Bloom (1964) suggests that a "conservative" estimate of the long-term effect of extreme environments may be about 20 IQ points. In supporting this statement, he notes that 20 points was the average difference for the three pairs of identical twins reared apart in the most dissimilar environments in the Newman, Freeman, and Holzinger study. He also cites a study by Sontag (1958) in which individuals changed as much as 20 points in what were termed as "favorable" and "unfavorable" environments. Burks (1928) suggested a similar figure for the effect of extreme environment.

These analyses of twin data indicate greater differences in intelligence test scores

between identical twins reared apart than Jensen acknowledges in his discussion; implied is a greater environmental contribution to the performance of even the most genetically similar individuals.

As indicated at the outset, this article could not cover all the issues raised in Jensen's lengthy discussion. Instead, it has dealt, to a greater or lesser degree, with some of the most salient problems raised, and has pointed out and offered corrections for a sampling of the errors and inconsistencies found.

In review of the areas covered, one central thread seems to emerge: that is, that Jenson's main omission is the picture of a complex and multifaceted environment, with which individuals interact in highly complicated and differentiated ways. Once that concept is firmly fixed, it would seem impossible to hold a simplistic view of the respective roles of heredity and environment in influencing intelligence test performance.

The burden of the discussion in the present article is the necessity for looking more closely at our environment in order better to understand the aspects which most impinge on individuals and influence their development, and in order to maximize those factors which exercise the most positive developmental influence and to minimize the most negatively acting ones. This is a tremendous task, and one which could well involve a large number of social and behavioral scientists. For not only will it be necessary to develop the requisite knowledge and understanding; it will also be necessary to feed the new knowledge past the organizational barriers and into the structures of society's institutions, most significantly the school system. Wilensky (1967) points up the kinds of difficulty to be expected:

> So often are accurate intelligence (i.e., information) estimates ignored—whether in the field or in the file of some subordinate department—that we might infer a general rule: the further we go from data collection to policy decision, the less knowledge and the more error—and indeed, standard treatments of intelligence imply some deterioration by stages. (p. 81)

Our society is in a very critical state of dysfunctioning. Unlike Rome, it could fall to a Carthage, either internal or external. The minds and knowledge of social scientists can play an enormous role in restructuring our social system as mediated through all human organisms. Through the socialization and education of children especially, it would seem that a significant degree of saliency could be reestablished between personal experience on the one hand, and on the other, social evolution founded in the gathering of knowledge and its correct and parsimonious utilization.

Unfortunately, Jensen's article, through its use by attorneys in some desegregation cases and by some legislators with respect to appropriations bills (aside from its overinterpretation in public media), has had a negative effect on social progress: less money for education cannot lead to better education; casting aside court desegregation decisions cannot lead to greater social equality.

As Dobzhansky said in the statement which was quoted more fully at the beginning of this paper: "Is a scientist accountable for misuses of his discoveries and utterances? He ought to be articulate enough at least to disown such misuses [1968, p. 129]."

Some years ago, I wrote an article on the concept of social courage, which I defined as an act "... taking place in a context of overt or covert social intimidation... [1959, p. 52]." The hypothesis was advanced that the manifestation of social courage would depend on the relationship between inner conviction (with respect to the issue around which the act would take place) and the punishment potential which the act would invoke. It would be in the social and scientific interest if Arthur Jensen would summon the social courage necessary to repudiate the positions which have been taken in his name; and to reexamine his thinking, reevaluate his sources of information, reassess his argument, and retract his genetic conclusions in the light of data about and understanding of environmental factors with which he was apparently not familiar at the time he wrote the article. In times of serious social crisis, when the barriers to social change are so enormous and when young people are venting such frustration, a senior social scientist's manifestation of the courage to reformulate a well-publicized opinion would be a positive example of the conquering of discomfort by the inner conviction of the necessity for scientific objectivity. It would be a positive act, too, because in the immense task which social scientists have with respect to our changing social structure, gifted experimentalists like Jensen can play important roles in generating new knowledge about the environment and the interactions individuals have with it.

References

American Institutes for Research. *A study of selected exemplary programs for the education of disadvantaged children.* Final report, Parts I and II. September, 1968, Palo Alto, Calif., Contract No. OEC-0-8-089013-3515 (010), Office of Education.

Blank, M. A methodology for fostering abstract thinking in deprived children. Paper presented at the conference on the Problems in the teaching of young children. Toronto, Ontario, March 1968.

Happenings on the Way Back to the Forum
MARTIN DEUTSCH

Bloom, B. S. Letter to the editor. *Harvard Educational Review*, 1969, 39, 419-421.

Bloom, B. S. *Stability and change in human characteristics*. New York: Wiley, 1964.

Bloom, B., Davis, A., & Hess, R. *Compensatory education for cultural deprivation*. New York: Holt, Rinehart, and Winston, 1965.

Burks, B. S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance. *Yearbook of the National Society for the Study of Education*, 1928, 27 (I), 219-316.

Burt, C. The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology*, 1966, 57, 137-153.

Caldwell, B. & Richmond, J. The children's center in Syracuse, New York. In L. L. Dittmann (Ed.), *Early child care*. New York: Atherton Press, 1968.

Cass, J. Race and intelligence. *Saturday Review*, 1969, 52 (2), 66-67.

Chein, I. On the nature of intelligence. *Journal of General Psychology*, 1945, 32, 111-126.

Coleman, J. S., et al. *Equality of Educational Opportunity*. Washington, D.C.: U. S. Dept. of Health, Education, & Welfare, 1966.

Covington, M. V. Stimulus deprivation as a function of social class membership. *Child Development*, 1967, 38 (2), 607.

Deutsch, C. Environment and perception. In M. Deutsch, I. Katz & A. Jensen (Eds.), *Social class, race, and psychological development*. New York: Holt, Rinehart, and Winston, 1968. Pp. 58-85.

Deutsch, C. P. & Silfen, C. Effects of a longitudinal enrichment program on language skills of disadvantaged children. Paper presented at the Society for Research in Child Development, Santa Monica, California, March, 1969.

Deutsch, M. Courage as a concept in social psychology. *Journal of Social Psychology*, 1961, 55, 49-58.

Deutsch, M. The disadvantaged child and the learning process. In A. H. Passow (Ed.), *Education in depressed areas*. New York: Teachers College Press, 1963. Pp. 163-179.

Deutsch, M. Social intervention and the malleability of the child. Fourth Annual School of Education Lecture, Cornell University, Ithaca, New York, May, 1965.

Deutsch, M. & associates. *The disadvantaged child*. New York: Basic Books, 1967.

Deutsch, M., Katz, I., & Jensen, A. R. (Eds.), *Social class, race, and psychological development*. Holt, Rinehart and Winston, 1968.

Dobzhansky, T. Genetics and the social sciences. In D. C. Glass (Ed.), *Genetics*. New York: The Rockefeller University Press and Russell Sage Foundation, 1968. Pp. 129-142.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and whites in the United States. *Psychological Bulletin*, 1960, 57, 361-402.

Dustman, R. E. & Beck, E. C. The visually evoked potential in twins. *Electroencephalography and Clinical Neurophysiology*, 1965, 19, 570-575.

Essen-Moller, E. Twin research and psychiatry. *Acta Psychiatrica Scandinavica*, 1963, 39, 65-77.

Feuerstein, R. & Shalom, H. Problems of assessment and evaluation of the mentally retarded and culturally deprived child and adolescent. The learning potential assessment device. Paper presented at the First Congress of the International Association for the

Scientific Study of Mental Deficiency. Montpellier, France, Symposium No. 8, September, 1967.

Feuerstein, R. *The role of social institutions and subsystems in the causation, prevention, and alleviation of retarded performance: A contribution to a dynamic approach.* Paper presented at the Peabody-NIMH Conference on Social-Cultural Aspects of Mental Retardation. Nashville, Tennessee, June, 1968.

Fried, M. The need to end the pseudoscientific investigation of race. In M. Mead, T. Dobzhansky, E. Tobach, & R. Light (Eds.), *Science and the concept of race.* New York: Columbia University Press, 1968. Pp. 122-131.

Fuller, J. L. & Thompson, W. R. *Behavior genetics.* New York: Wiley, 1966.

Gibson, E. J. Perceptual development. In H. W. Stevenson, J. Kagan and C. Spiker (Eds.), *Child Psychology: The 62nd Yearbook of the National Society for the Study of Education,* Part I. Chicago: National Society for the Study of Education, 1963. Pp. 144-195.

Gibson, J. J. & Gibson, E. J. Perceptual learning: Differentiation or enrichment? *Psychological Review,* 1955, 62, 32-41.

Glasman, L. D. A social-class comparison of conceptual processes in children's free recall. Unpublished doctoral dissertation, University of California, 1968. Cited by A. Jensen, How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969 39, 1-123.

Gottesman, I. I. Genetic aspects of intelligent behavior. In N. R. Ellis (Ed.), *Handbook of mental deficiency.* New York: McGraw-Hill, 1963. Pp. 253-296.

Gottesman, I. I. Beyond the fringe personality and psychopathology. In D. Glass (Ed.), *Genetics.* New York: Rockefeller University Press and Russell Sage Foundation, 1968. Pp. 59-68. (a)

Gottesman, I. I. Biogenetics of race and class. In M. Deutsch, I. Katz, & A. Jensen (Eds.), *Social class, race, and psychological development.* New York: Holt, Rinehart, and Winston, 1968. Pp. 11-51. (b)

Guilford, J. P. Three faces of intellect. *American Psychologist,* 1959, 14, 469-479.

Guilford, J. P. *The nature of human intelligance.* New York: McGraw-Hill, 1967.

Haller, M. Social science and genetics: A historical perspective. In D. C. Glass (Ed.), *Genetics.* New York: Rockefeller University Press and Russell Sage Foundation, 1968. Pp. 215-225.

Hebb, D. O. The motivating effects of exteroceptive stimulation. *American Psychologist,* 1958, 13, 109.

Hefner, R. Personal communication (typescript), 1969.

Hilgard, E. R. & Atkinson, R. C. *Introduction to psychology* (4th ed.). New York: Harcourt, Brace, and World, 1967.

Hirsch, J. Behavior-genetic analysis and the study of man. In M. Mead, T. Dobzhansky, E. Tobach, & R. Light (Ed.), *Science and the concept of race.* New York: Columbia University Press, 1968. Pp. 37-48.

Jencks, C. Some natural experiments in compensatory education. Unpublished manuscript, April, 1969.

Jensen, A. R. Social class and verbal learning. In M. Deutsch, I. Katz, & A. Jensen (Eds.), *Social class, race, and psychological development.* New York: Holt, Rinehart, and Winston, 1968. Pp. 115-174.

Jensen, A. R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1-123.

Juel-Nielsen, N. Individual and environment. A psychiatric-psychological investigation of monozygous twins reared apart. *Acta Psychiatrica Scandinavica*, 1964 (Monog. suppl. 183).

Katz, I. Factors influencing Negro performance in the desegregated school. In M. Deutsch, I. Katz, & A. R. Jensen (Eds.), *Social class, race, and psychological development*. New York: Holt, Rinehart, and Winston, 1968. Pp. 254-289.

Katz, I. Review of evidence relating to effects of desegregation on the intellectual performance of Negroes. *American Psychologist*, 1964, 19, 381-399.

Klineberg, O. *Negro intelligence and selective migration*. New York: Columbia University Press, 1935. (a)

Klineberg, O. *Race differences*. New York: Harper, 1935. (b)

Klineberg, O. Negro-white differences in intelligence test performance: A new look at an old problem. *American Psychologist*, 1963, 18 (4), 198-203.

Knobloch, H., & Pasamanick, B. Prediction from the assessment of neuromotor and intellectual status in infancy. In J. Zubin & G. Jervis (Eds.), *Psychopathology of mental development*. New York: Grune & Stratton, 1967. Pp. 378-400.

Kozol, J. *Death at an early age*. Boston: Houghton Mifflin, 1967.

Kuttner, R. E. Letters to and from the editor. *Perspectives in Biology and Medicine*, 1968, 17, 707-709. Cited in A. Jensen, How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1-123.

Lee, E. S. Negro intelligence and selective migration: A Philadelphia test of the Klineberg hypothesis. *American Sociological Review*, 1951, 16, 227-233.

Mental tests for 10 million Americans—What they show. *U. S. News and World Report*, 1966, 61 (1b), 78-80. Cited by A. Jensen, How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 1969, 39, 1-123.

Montagu, A. *Prenatal influences*. Springfield, Illinois: Charles Thomas, 1967.

Nelson, G. K., & Dean, R. F. A. The electroencephalogram in African children: Effects of kwashiorkor and a note on the newborn. *Bulletin of the World Health Organization*, 1959, 21, 779-782.

Newman, H. H., Freeman, F. N., & Holzinger, K. J. *Twins: A study of heredity and environment*. Chicago: University of Chicago Press, 1937.

Pasamanick, B. A tract for the times: Some sociobiologic aspects of science, race, and racism. *American Journal of Orthopsychiatry*, 1969, 39 (1), 7-15.

Pettigrew, T. Negro American intelligence: A new look at an old controversy. *Journal of Negro Education*, 1964, 33, 6-25. (a)

Pettigrew, T. *A profile of the Negro American*. Princeton, N. J.: Van Nostrand, 1964. (b)

Piaget, J. *The origins of intelligence in children*. New York: International Universities Press, 1952.

Powledge, F. *To change a child*. Chicago: Anti-Defamation League, 1967.

Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom*. New York: Holt, Rinehart, and Winston, 1968.

Scientists bar genetic study of the Negro, *New York Post,* April 30, 1969 (byline, Victor Cohn), p. 30.

Scott, J. P., & Fuller, J. L. *Genetics and the social behavior of the dog.* Chicago: University of Chicago Press, 1965.

Scott, J. P. Discussion: Race as a concept. In M. Mead, T. Dobzhansky, E. Tobach, & R. Light (Eds.), *Science and the concept of race.* New York: Columbia University Press, 1968. Pp. 59-68.

Shields, J. *Monozygotic twins brought up apart and brought up together.* London: Oxford University Press, 1962.

Shields, J. & Gottesman, I. I. Age at separation and IQ differences in identical twins reared apart. Unpublished manuscript, 1965. Cited by I. I. Gottesman. Biogenetics of race and class. In M. Deutsch, I. Katz & A. Jensen (Eds.), *Social class, race and psychological development.* New York: Holt, Rinehart, & Winston, 1968. Pp. 11-51.

Shockley proposal tabled. *Science,* 1969, 164, 652.

Smilansky, S. Progress report on a program to demonstrate ways of using a year of kindergarten to promote cognitive abilities, impart basic information and modify attitudes which are essential for scholastic success of culturally deprived children in their first two years of school. Jerusalem: Henrietta Szold Institute, 1964.

Sontag, L., Baker, C. E., & Nelson, V. Mental growth and personality: A longitudinal study. *Monograph of the Society for Research in Child Development,* **23**, No. 2, 1-143.

Spearman, C. *The nature of intelligence and the principles of cognition.* London: Macmillan, 1923.

Spuhler, J. N., & Lindzey, G. Racial differences in behavior. In J. Hirsch (Ed.), *Behavior-genetic analysis.* New York: McGraw-Hill, 1967. Pp. 366-414.

Stoddard, G. D. *The meaning of intelligence.* New York: Macmillan, 1943.

Stone, L. J. & Church, J. *Childhood and adolescence* (2nd ed.). New York: Random House, 1968.

Stott, D. H. Interaction of heredity and environment in regard to measured intelligence. *British Journal of Educational Psychology,* 1960, **30**, 95-102.

Stott, D. H. *Studies of troublesome children.* New York: Humanities Press, 1966.

Tanser, H. A. *The settlement of Negroes in Kent County, Ontario.* Chatham, Ontario: Shepard Publishing Co., 1939. Cited by E. R. Hilgard and C. Atkinson, *Introduction to psychology.* New York: Harcourt, Brace, and World, 1967.

Thurstone, L. L. *Primary mental abilities.* Chicago: University of Chicago Press, 1938.

U. S. Commission on Civil Rights. *Racial isolation in the public schools.* Vol. 1. Washington, D. C.: U. S. Government Printing Office, 1967.

Using speed of brain waves to test IQ. *Medical World News,* 1968, **9**, 26. Cited in A. Jensen, How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969, **39**, 1-123.

Vandenberg, S. G. Contributions of twin research to psychology. *Psychological Bulletin,* 1966, **66**, 327-352.

Vandenberg, S. G. The nature and nurture of intelligence. In D. C. Glass (Ed.), *Genetics.* New York: Rockefeller University Press and Russell Sage Foundation, 1968. Pp. 3-58.

Von Senden, M. Raum und Gestaltauffassing bei operienten Blindgeborener von und nach der Operation. Leipzig-Barth, 1932. Cited by D. O. Hebb, *The organization of behavior.* New York: Wiley, 1949. Pp. 28-31.
Wheeler, L. R. A comparative study of the intelligence of East Tennessee Mountain Children. *Journal of Educational Psychology,* 1942, 33, 321-334.
Wilensky, H. *Organizational intelligence.* New York: Basic Books, 1967.
Woodworth, R. S. Heredity and environment: A critical survey of recently published materials on twins and foster children. *Social Science Research Council Bulletin,* 1941, 47.

The Politics of Pronouncement: Notes on Publishing in the Social Sciences

THOMAS J. COTTLE
Harvard University

Focussing on the estranged reaction of individuals to scholarly writings about their ethnic groups, Thomas Cottle explores a network of political implications surrounding publishing in the social sciences. This network extends from published content through the act of publishing itself. He describes the interactions of political motives, conceptions of the university, communication media, and the public to convey a sense of the political ramifications of publishing in the social sciences.

The journal closed on her lap. Erlene Menter lay back in her chair, her legs stretched out, her eyes wide open, looking high into the corner of the small room as if there might be something up there for her to read. "They sure do write hard English," she finally said. "By the time you get around to reading the thing, then understanding it, it sure seems as though you've been on a long, long, long trip. Now you tell me, do you really think any scientist or what you call doctor, is going to understand what my life is really like? Are you going to tell me that these fellows from Harvard with all their books and schooling are going to

Harvard Educational Review Vol. 39 No. 3 Summer 1969

have the slightest eye full of this room, and the kids? Why, if they saw this mess they'd up and leave in a minute, and you know they would. The students, well, they're different. They come around here with their jabbering and all, all excited, ready to make trouble; they're gonna make war on the world if anybody'd give them half a chance. They're like you. They got these little pieces of paper for me to read about housing, and bussing and welfare laws. Some of it makes good sense, like they know those rules about welfare. Don't you kid yourself for a minute. They know those rules and most of them aren't even lawyers."

To read the essays, documented as they were with the figures of certainty and authenticity, was to be invaded, molested, as it were, swallowed up by the rough-grained pictures of professionals who, in a funny way, had no business being there at all. It was not the explicit political position of these articles that hurt as much as the sense that sacred proprieties had been ignored. It was as though curfew laws had been violated by the very men who had established them in the first place.

Evidently, Kathleen Cavanaugh had been waiting eagerly for my arrival. Though I was on time, she acted as if I'd kept her waiting three hours for our appointment. She was steaming mad, more upset than I'd ever seen her, even more so than the time we had spoken of Robert Kennedy's death. That had been several months after the assassination. Her self-proclaimed period of mourning concluded, a time of uncontrollable anger had overtaken her. She just didn't know what to do with all that anger, and my suggestion that we take a walk didn't seem to help.

But on this more recent occasion, the seventy-three-year-old widow was burning. She was all but ready to pounce on me when I knocked. Usually I would see her through the glass, descending the long staircase, straightening her skirt, just as she would reach the inside of the door. This day, however, she was peering out, surveying the street as if forewarned of an impending accident.

"I've read it. That pack of lies. Who in the name of You Know Who gave you something like that to give to me? Why the nerve of those people. Say, Thomas, they're not friends of yours, are they? That's good. Think of that. Here I am shooting my mouth off and maybe saying things about your friends. But they're not at Harvard? They couldn't be? That's good. A man's got to be real careful saying those things. Why there's not a morsel of truth in that. I've been alive on this good earth almost seventy-four years. Even a woman doesn't go to high school learns something in that time. You have to, raising your children and all. I

showed this to the boys upstairs. You know I told you I rent the upstairs to those Harvard boys. Well, they agreed with me. One said he's read something just the other day which says just the opposite. Now how's that possible? Either these are real facts or they're make believe, and scientists don't make things up, at least the ones on TV I see don't make things up." We laughed at that.

Two women, anyway, were unable to discover in these pages the truly valued aspects of their lives. Supposedly they were to see themselves on the neatly printed page, but all that emerged were tragic distortions of themselves and of their worlds. Somewhere the essence of their lives had been lost in the waves of categories, data analyses, and discussions of findings. Suddenly there had emerged on the clean, white paper mere content, samples, not of people whom they did not know nor barely recognized, but of themselves.

The trip to Hannah Brachman's was always interesting. Travelling down Blue Hill Avenue, the most direct route to her house, revealed a panorama of Boston's social history. How many students had come to this area to make their own studies, and then described their impressions of the soul food stores alongside the Kosher butcher shops or the Mogen Davids adorning the cement fronts of record and barber shops. Everyone knows the "story" of blacks moving in, Jews moving out, and the exodus of the young and the affluent to Brookline and Newton.

Mrs. Brachman was always waiting for me, some food prepared, a neighbor's child reading in the kitchen, eager to make friends with someone from a University. Our discussions usually centered on her feelings about her family, Jewish writers and scholars, events in New York, Jerusalem, and at Brandeis. Always she would have words of praise for the president of Brandeis and more harsh words for blacks occupying buildings and claiming that the University's name should be changed. Mrs. Brachman's loyalties were coming out more strongly. "Who helped us?" she would ask rhetorically. "Who marched down Fifth Avenue or sat in buildings or made revolution for us? They still don't do anything for us."

As a favor, Mrs. Brachman read a couple of articles written by social scientists on the activities of Jews in the New Left and the rather significant position of power they seemed to have attained. We both agreed that the pieces were written from sympathetic viewpoints, the authors presumably remaining as objective as possible. The data, I had thought, were well collected, thoughtfully analyzed, and presented without bias. "You can't argue with those numbers. It's hard to fight that. It seems pretty obvious." Soon her eyes moved away from the pages of the reprints, now wrinkled and torn. "You know, not a lot of people know it, but the

Jews have done a lot for this country. When you stop to think of all the doctors and lawyers, all the professors, it's really something. It's really some accomplishment. And look at Israel. Is that not something extraordinary? The fears, the wars. What these people have suffered. From one war right into that trouble. What's going to happen? What's going to happen?"

Where in this was the reality our students wish us to discover as they yank us into the world hoping that our observations might be more accurate, our recommendations for change more influential? How are we able to differentiate our research intentions from our policy-making intentions? And how can we separate our desire to make science from a publisher's or reader's desire to make politics?

"What about the article, Mrs. Menter?"

"Oh yes. Well, I don't care what he says here. I know about that Coleman report and this report and that report. You don't need to tell Negroes about that stuff. That's white man's words for white man's ears. When the Negro professors start writing things, you'll see a whole different picture. You go out and bring me some of their work and you'll get a different picture. You'll get a very different picture."

Erlene Menter knew full well the contents of the eight page article I had asked her to read. When all the grammar, paragraphs, and data had been pushed aside, she saw a terrifying message, naked and bleeding. She had read eight pages about black children growing up in ghettos, about absent males, the occurrence of incest and the impact of all this on children and on a race of really not so many people who were struggling to find a pattern that might simultaneously knit them together and then, bounce them all up, upward to where they wouldn't receive such devastating rebukes or, for that matter, such perplexing triumphs. The message she got, was that when this scientist ran his figures and numbers through a computing machine, it came out, as she said, "that the Negroes aren't getting anywhere in particular, too fast."

Debates on the possibility of "value-free" social science are becoming increasingly rare. Some social scientists believe that they can make so-called "value-free" contributions to fact and theory. Others are sure that this freedom from bias can never be achieved, if due only to the more subtle implications of the very act of publishing from a position within a university.

As effortful as each day had become in the eighteen years since Francis Cavanaugh died of a heart attack in the house where she still lives, Kathleen Cava-

naugh fulfilled her promise to me by reading twenty rather trying pages on the value and belief systems of working-class Catholic families. I picked the piece especially because we had spoken of such matters before, and because it seemed to me, anyway, that the authors had captured without obvious distortion the lives of people "sociologically similar" to Mrs. Cavanaugh. I had thought she would immediately read of a familiar world, accurately presented. Diligently and methodically she had read the assignment, even taking rather copious notes on the inside of the telephone book in that delicate thin-line handwriting of hers.

"Are they teaching this kind of stuff at your school, because if they are you could sure do a good thing for these students by telling each and every single one just what we do believe. And don't you let them get away with this. I'll bet you those professors never did speak with any of those people they write about. No one talks like that, unless he's composing something, like a story or poem."

Kathleen Cavanaugh was profoundly upset. The article had portrayed something insidious. Undeniably, she had felt betrayed. It was as though her pride had been extinguished, her very soul invaded and found dry and hollow. She had learned more, she would say later on, from television even though it too "favors what the rich people have to say and think. And buy." By what had seemed to me to be an insightful, penetrating glance at a community's social life, Mrs. Cavanaugh had been shot down, right in her steps.

The implications of disseminating the research findings she had read had not been lost on Mrs. Brachman, either.

"Do you know some of the things the students would like to see changed in society?"

"Everyone knows. Even in Washington they know. They don't like this war. Who can like that war? Can you imagine this business with the boat, this Pueblo business? The kids don't like that, do they? They think it's unfair? I can't blame them. Why are we fighting and spying and killing? Every night on the television that's all you see. Tell me, is it true what I read, Jews are really running these college things? Maybe someone should tell them it doesn't look so good. Do you think it's good for people to read such things, even though they say it's true, you know who I mean? A lot of people read articles like this, don't they?"

It seems almost impossible to publish a report that represents no political bias or implies no political action. Whatever our intention, whatever our assumptions of how "value-free" our research can be, the implications stay with us. Even

Politics of Pronouncement
THOMAS J. COTTLE

with our modest intentions to "advance science or knowledge," the popular media and its readers stand ready to greet the applications or the political implications. What, in short, they ask, are the products or profits and the statements of appropriate action to be found in these writings? What can I take and use of this? How can it be reduced to the solid, true laws of human nature that these scholars are, after all, supposed to be discovering?

To these questions, social scientists respond with troubled ambivalence. Pressures from many people have been put on academicians to derive with certainty the state of human nature and programs for the upgrading of everyone. Working against this, naturally, are the "limitations of the art" as well as, perhaps, a primordial reluctance to explain mankind, to explain so much variance that futures become predictable, presents explicable, pasts logical, and certainty guaranteed. There just may be a primitive sense in each of us that will forever prevent a total explanation or perfect experiment. Yet if such a sense exists, it may not be tolerated by audiences demanding exactitude in diagnosis and treatment.

Still, we do little to convey to these audiences the tentativeness and possible inaccuracy of our statements. There are those of us who qualify their televised pronouncements with "we know very little," or "our science is so young," only to proceed to deadly pontification. Others advance the most recently achieved knowledge while ignoring the attendant responsibility of their published words.

Recently, a young social psychologist bemoaned the overnight success of his first book. He had received letters from everywhere, even from soldiers in Viet Nam, asking him whether they could take his tests and undergo his experiments, which somehow were supposed to better their lives. What shocked the author, really, was the way people "could just take over my book and do with it whatever they wanted." No longer was he in control. From even a cursory reading they had come away with political and social strands he himself barely recognized. Where he had used data to reinforce hunches, they had clutched that data as proof of the book's "real" message. They had skipped over the pages where conceptualizations were embellished and had rushed instead to the meaty parts from which they might take something for themselves. Now they begged him to let them be a part of his grand scheme for change and success. There seemed to be nothing in their reactions suggesting an appreciation for any intellectual contribution. "They read that book as though it were a manual on how to ice skate."

His book was taken as a manual because in part the media of popular communication cannot always tolerate messages of what intellectuals think about, work with, or, indeed, play with. Popular media cannot always permit the luxury

of theoretical reasoning or development, nor can they spend time dealing with contributions to the history of theory when there are hard, cold facts to be gotten out and publicized. Moreover, there must be a splash, a glimmer, a scintillating explosion in each and every published pronouncement or it won't "catch on." There must be something that one can hold in his hand, a "fistful of reality," as Sartre said.

The conflicting needs of scientists as against those of their publishers, readers and, increasingly, the students, make it progressively more difficult to "get away with" pure and simple contributions to theory and methodology. Despite the many failures and the flood of contradictory books and reports, much of the public remains loyal to the belief that social scientists are experts, suppliers of the right kind of knowledge. In a word, their expertise renders them "solutionists." Their ideas cannot stay as ideas, but must be translated into facts and answers. As speedily as these ideas pass from the page to the eye, they lose their tentativeness and "hunchiness" and become certainty as well as plans for action.

Erlene Menter was laughing again, sitting up straight and pushing the journal back across the table. As it moved, she rotated it slightly, the letters now right side up for her. "Nice colors they use," she said, staring at the cover and fondling its smoothness as though the outline of each letter might stand up just high enough so that she could touch it, then read it with her eyes closed. She let the pages riffle gently along the tips of her fingers, then a few times more. "The paper's nice too. Not like the newspaper."

The article had said as much through its authoritative, bookish appearance as it had through the statements on its pages.

Just as what we study represents a very real system of values, so too do the "products" of our studies perpetrate these values and hold them up as some ideal, however temporary. A popular conception holds that in science, publication implies certainty. Clearly, too much certainty is taken for granted. Among most readers, even editors, scientists simply cannot play with ideas. Tentativeness and unsureness cannot be accepted from them. Maybe that's why correlations too often emerge as causation and why summaries of findings get publicized as incontestible facts.

For Hannah Brachman, a mythic tradition of intellectualism and achievement, spirit and honor along with suffering, welled up within the soul she chose to share with millions and millions of people. The two studies she had read were bad press; they could not be denied, shoved aside or forgotten. Scientists teach

facts, and the facts they had taught her were that Jewish boys and girls were being disruptive, causing problems, getting themselves into serious trouble and going to jail. For her, science is facts, undeniable, incontrovertible facts. "When a man with such education, such erudition speaks, he knows what he's talking about. Maybe I'd like to disagree. To tell you the truth I wish he hadn't written this. Or maybe I wish you hadn't brought it to me. But that's the world. That's the world. It seems a shame."

Some people, naturally, have "adopted" the findings of social sciences and found them valuable for their work and for their lives. But the day is not yet here when the "public" fully appreciates the playfulness of ideas or the fun and excitement of knowledge. Not enough people yet understand the little boy or girl, free from everything and everyone, alone in his room, deeply engrossed in a task only angels dare understand.

Surely there still exists the popular conception of the professor as the man who is "only" playing. This is the notion that speaks to his lack of any tangible product or of "an honest day's work" and concludes that the professor remains as childishly occupied as the children he teaches.

This is hardly the same view as that held by academics about the playfulness of ideas inside the academy. The evolution of intellectualism, just as the development of cognitive abilities in the child, brings cultures to the point where ideas almost stand by themselves, unencumbered by political association with some greater shared reality. Indeed, the highest form of thought permits both the capacity to imagine the impossible or unreal and the capacity to play with ideas, to work with and sculpt them, even if the final product fails to yield anything but joy.

Now, as students argue louder than ever before, only the very elite can still afford the luxury of such playfulness and tentativeness. Only the elite can dare consider "intellectual contributions" sufficient. And yet they must be made. Many students have joined the public in crying for political products and not playfulness. And so, faculty members now fear the end of purely "academic days" as they struggle to defend themselves against what they feel to be an onslaught of anti-intellectualism, anti-rationalism, anti-objectivism and anti-science led by, of all people, their very disciples and apprentices.

"Think of the money spent trying to figure out what's happening in these neighborhoods. That other book you had JoAnne (her daughter) read was all about black folks in Baltimore and Washington. Think of that. They go all the

way to Baltimore just to look into their homes when they could come right here. They're all welcome right here. You tell them if they want to make some of their studies, they should come and see me. I'll tell them stories they can write ten books about, fifty books about if their hands don't get tired and those machines of theirs don't die."

"I think maybe those studies were done by people who lived in Baltimore and Washington."

"Maybe so. I thought JoAnne said something about going all the way down there to make their studies. You don't hear anything about this neighborhood, 'cepting that there's trouble with the welfare boards and those... Man, they've got a collection of people working for them, you wouldn't believe your eyes. Not too many of your Harvard folks, I'll bet."

Two sorts of familiar political spectrums, really, have emerged: the "horizontal" scale to the left and right of moderate and the "vertical" spectrum about which our students are teaching us. The "vertical" scale extends from elite privilege to disenfranchisement. Coming from a generation of objectivity, students have long advocated total awareness of this spectrum but now demand direct participation in the lives of disenfranchised and oppressed people.

The university model of detachment and non-involvement was seriously shaken by the Civil Rights movement of the 1950's. The initial student involvement in the lives of southern Negroes lead to their emphasis on political intervention and on becoming implicated. Sit-ins quickly turned to voter registration and redistricting campaigns. But the intellectuals remained a step behind, some reporting on the events in the North and South, many banging out research documents investigating the parents, grandparents, school problems, and generalized psychopathologies of student workers. Nevertheless, the result for many scholars was a violent shift from playfulness and sovereign academic goals to a politicization of their research in a way that would, as they say, help mankind. At the very least, this new breed of social scientists was thinking about the concrete products of their enterprises and the implications these products might have in the political arena. For them, academia was a necessary home and tentativeness a necessary constraint, but not a way of distancing themselves from a population for whom they cared, and at times, for whom they grieved. They wanted and needed to be in touch.

JoAnne Menter sat cross-legged on the floor; two friends slouched on the sofa listening to her read sections from a book. As she recited certain passages care-

fully marked by her earlier, they all screamed with laughter, bouncing up and down from their scattered positions on the floor and furniture. JoAnne would start another passage, and they would cackle and jabber. "You better believe it, baby. This cat sets up right there on the corner . . . ." And they would roar. I couldn't help laugh myself. Erlene, working about as though she weren't paying the four of us any attention, showed by an occasional glance that she would just as soon send all of us maniacs to some institution. But she too understood.

As Civil Rights movements and now Vietnam have exploded all students out of the narcissistic pleasantries of psychological reasoning into the more profane acreage of sociology and political or policy sciences, an implicit hope has developed that social science will not only be "relevant," but chock full of policy implications. Some scientists have responded directly by sitting on government commissions; others respond less directly by consulting, a tenuous process in which, almost rheostatically, they may control the amount of their commitment and involvement. Despite a prevalent anti-intellectualism, the contention persists, an almost spiritual contention, that knowledge is power and with it no limits need be set. Surely if we can get to the moon, we can get to inner cities, suburbs and Appalachia, southern Texas and Florida. Nonetheless, while some scientists dive headlong into the explicit politics of their research, others seek to wiggle out of the politics of certainty. Not wishing to participate in intellectualized politics of confrontation they strive to keep their distance, if not necessarily their "disinterestedness." They too, however, have become aware of the political implications of their work and are, perhaps, becoming aware of their place in that vertical spectrum.

Political self-consciousness now has grown to the point where we recognize and confess to the more obvious implications of our printed statements and of our acts of publishing. No one needs to be told, for example, that in such areas as "race relations" debates in the literature or disagreements over interpretations of data or over the assessment of methodological steps have more than a latent political impact. This we know. We have all shuddered a bit in the last weeks. But are we always able or willing to shudder at research pronouncements that do not seem to us to be so "touchy"—for example, studies of working-class Catholics or the activities of Jews in the New Left? Do we shudder, in fact, from the politics embedded in gigantic volumes of theoretical scripture? Are we not now reading theoretical expositions in part for the politics they might bequeathe? Are we, because of the people who read us and publish us, becoming

aware of touchiness and relevance, discreteness and ethicality, against a backdrop which heretofore has been infrequently used? And, from all of these issues, may we ever again claim objectivity, or freedom from politics and from elitism?

"Did the articles remind you of anything, Mrs. Brachman?"

"The articles. The articles tell me two professors, two just like you, are telling me that this college business is being run by Jews. Jewish boys and Jewish girls. This part I can't figure out at all. What business is it of the girls? If they're going to get into trouble, at least it should be the boys. What do these girls know? They're so young. They're so small. Aren't they interested in ... in ... in growing up, with homes, with wives, with husbands, with children! What's it coming to with Negroes fighting with the police, with boys and girls in the schools fighting with their teachers? They should go without a little bit, they'd see how you fight with policemen!"

Some scholars now notice the more subtle political strains which silently contribute to the kinds of research topics chosen by scientists and the kinds of research "acceptable" to the publishing and reading public. While scientists need not think in these terms, no one can doubt the fads and ritualized sources of inquiry generated and perpetuated in the social sciences. In fact, they have been studied. But fads themselves are steeped in the juices of politics. Thus, journals and magazines propagate politics when they select topics or writers whose extravagant or subtle polemics go in the proper direction. One finds, for example, mountains of articles on inner city schools but nary a molehill on the problems in suburban high schools, apart from drugs, long hair, and dress codes. There is more than one can read on working-class patterns but surprisingly little on the upper middle class. As someone said, a bit severely, perhaps, to study the working class is sociology, to study the rich, exposé.

Kathleen Cavanaugh pounded her fist on the open pages: "They're just not not going to pin me down that easily." JoAnne Menter, too, had rebuked the characterization of "her people." Her laughter hardly masked the poignancy of the book and her desperate attempts to climb out from under the shackles of categories, divisions of populations, and conceptualizations made by some "smart guy who thinks he knows us just 'cause he's been to school longer than us." As much as in their fight to stay abreast of groups, collectivities, cultures, and hordes of people they could hardly imagine—even after attending a giant rally for black people only—Erlene Menter and her daughter fought hardest of all to maintain a

Politics of Pronouncement
THOMAS J. COTTLE

single stretch of their own being and their own singular identities. "Before anything else," JoAnne would say, "I am always me. Somebody told me that God exists in each of us and that we should be proud just to be ourselves. So, I'm going to be me, and if people don't like it. . . ."

Politicized students have managed to convince many academics that even if they shy away from research that has explicit policy implications or from polemical pieces which unequivocally indicate their political persuasion, the very actions of research and writing can be deemed elitist. Our concern for the working class or the blacks is lovely to behold, they argue, but when we offer our ideas as weighted as they are by our proclaimed status they cannot help but be blistered by the dispositions of our enterprises and by the politics of our lives and life styles. Like air bubbles, politics has been pumped into the research of people who have worked diligently to make sure none would be found.

Indeed, the day may never come when students succeed in pushing all scholars into what they call reality. By reality they mean not only that intellectuals should become involved, engaged, politicized, but that they should be aware of the political electricity that illuminates their writing and acts to legitimate their cause and freedom. Students argue that universities cannot remain isolated unless people like Kathleen Cavanaugh or Erlene Menter have a place; and until that time they cannot condone the political elitism of studies and offerings which, in their very prose, protect and distance us from those we study. How often, they ask, would we admit to knowing our "subjects" and "respondents"? How often do we consider the pretense at objectivity which removes us from the world in which we observe and write and think? How often, they ask, do we take seriously the political positions from which our writings unwittingly take shape and from which policy statements ultimately are drawn? And always they throw that word, "elitism," at us, in an effort to extinguish our habits of playfulness and immodest indifference. They want us out of our offices and "into the world." Many of them want our voices to come together in what they call a "new politics."

One of many justifiable statements heard in rejection of assuming political stances or of doing explicitly policy-oriented research is that these actions too easily lead to governmental or societal restraints on research topics and operations. There is much to fear if research is taken over by the constraints imposed by any interest group.

A paradoxical result of current student focus on the politics of their professors is that one utterance can forever—publicly and inaccurately—nail a faculty mem-

ber to one political position just as he dares to step out of his office and into the realities of a stratified society or into what we call "the field." We have yet to realize fully that the profound political implications of our work do not lie on the left-right political spectrum in which our audience might stereotype us, but in our witting or unwitting participation in that other spectrum which contains poverty, racism, disenfranchisement, and oppression.

The most telling sign of this may be that Mrs. Cavanaugh and Mrs. Menter cannot find themselves in the articles they read. They cannot get the picture moved around so that it includes them. At least this is what they say. For it also may be true that the studies have found them too accurately, too penetratingly and, as sensitive human beings, they must recoil from the unintended stabs and stereotyping of these portraits.

Their reactions, therefore, might be what some call "denial." But if it is denial, it may have something to do with the fact that the studies' portraits bring them nothing more than reading materials from a teacher. Their reactions may well be natural protection against a hope that more might come, that something might happen. For while we in our debates over interpretations of data may take time out for reanalysis, reevaluation, or even for play, they dare not leave the apartments where their children will be raised, nor the houses where their husbands died, for even a moment of truly fresh air.

The self-insulated separation, the lack of sensitivity to the vast and subtle political implications of our publications in part come down to our not hearing the quiet phrases and the ritualized language forms which too often go unnoticed. *When I left Mrs. Brachman for the last time she walked me to the door of the apartment, always so neat and open to guests and family. She looked me squarely in the eyes without shame and without defiance: "When you're done with your work and you have a little time on your hands, you'll go with your wife and you'll get a haircut, and maybe you'll find some time to come back and we'll talk a little. The three of us. O.K.?"*

*Critique of Hereditarian Accounts of "Intelligence" and Contrary Findings: A Reply to Jensen**

F. S. FEHR
University of Illinois

The confounding of environmental and hereditary influences is a considerable problem in estimating heritability from twin studies. Fred S. Fehr discusses this problem and suggests two ways of calculating heritability which separate these influences more cleanly than the formula commonly used. The importance of heritability in the determination of intelligence is considerably less than suggested by Jensen when the effects of environmental variables are thus more adequately controlled.

"Impenetrability! That's what I say."
"Would you tell me, please," said Alice, "what that means?" .... "I meant by 'impenetrability'," said Humpty Dumpty, "that we've had enough of that subject, and it would be just as well if you'd mention what you mean to do next, as I suppose you don't mean to stop here all the rest of your life." (Lewis Carroll, *Through the Looking Glass*, 1873, Chapter 6)

The purpose of this article is to call attention to some apparently disregarded and/or overlooked findings in the frequently quoted and classic study of Newman,

* Supported by Grant No. USPH NB 07346 from NIMH.

Harvard Educational Review Vol. 39 No. 3 Summer 1969

Freeman, and Holzinger (1937), and to provide findings derived from the equally noted results of a study by Shields (1962). These findings are indirectly contrary to Jensen's conclusions (1969) regarding the relationship of the heritability of intelligence[1] and social class; they are specifically contrary to the underlying and necessary assumption that individual differences in intelligence can largely (if not almost entirely) be accounted for by hereditary factors.

Confounding of Variables

Before examining these data, several comments and points are in order. One frequently considered objection to correlational studies purporting to demonstrate the importance of heritable, as compared to environmental, factors is that both factors are usually confounded in such efforts. This is particularly relevant in comparisons of the estimated intelligence of monozygotic (MZ) and dizygotic (DZ) twins when the twin pairs live from birth in the same family environment. Thus, the greater resemblance of the estimated intelligence of MZ twins as compared to DZ twins may be due to the environment being more similar for the MZ pairs. That is, the marked similarity of appearance of identical twins causes other people to treat them alike and confuse them with one another. They are often dressed alike and treated as a unit by their family and friends. Wilson (1934) concluded that both fraternal and identical twins share a more similar environment than siblings, and identical twins a more similar environment than fraternal twins. The finding of greater similarity of environments for fraternal twins than for siblings has also been supported by Herrman and Hogben (1932).

Seemingly, the significance of this criticism has generally been little appreciated. However, when one considers the minimal twin pair difference obtained in deriving the percent variance which is purported to be an estimate of heritability, the issue may be placed in more adequate perspective for the data to be subsequently presented here. Specifically, the mean difference between DZ twins reared together has been reported to be 9.9 points and for MZ twins 5.9 points on the Binet IQ test (Newman, Freeman, and Holzinger, 1937). Coefficients of heritability have been estimated from 65 to 80 percent from these same data. The point to be emphasized,

[1] Tests of intelligence in the present context are referred to in the spirit of Hempel's thesis (1965) of liberalized operationism, that is, that tests as operations are presently the most reliable indicators of intelligence and are not necessarily any finalized valid measure of what some authors might claim intelligence to consist of. Thus, intelligence in this frame of reference is indicated by widely utilized test measures.

if not already apparent, is that such minimal mean differences between MZ and DZ twin pairs (4 IQ points in this case) can be accounted for equally well by environmental factors. To reiterate, mean differences between twin pairs of 4 IQ points can be accounted for by the fact that MZ twins share a more similar environment than do DZ twins. Studies of twins separated in early life and reared apart would seem to provide an adequate basis to answer the aforementioned criticism. But an examination of these studies reveals both methodological difficulties and contradictory findings.

Methods of Analysis

As suggested by Jensen (1969), studies of twins reared apart are not only the simplest means of estimating heritability coefficients, but also a methodologically more adequate means of limiting the confounding of environmental and genetic factors. According to Jensen "all they (separated MZ twins) have in common are their genes," a point which will receive attention shortly. As Jensen also notes, there are relatively few studies of separated twins. The ones that are most prominently reported in the literature are those of Newman, Freeman, and Holzinger (1937), Shields (1962), and Burt (1966). With some variations, several formulas have been initiated and/or derived from the Newman, Freeman, and Holzinger study for the estimation of heredity and environment. These include comparisons of (1) MZ and DZ twin pairs reared in the same environment, (2) MZ reared in the same environment and MZ pairs reared in separate environments, and (3) MZ and DZ twins reared in separate environments. These will be elaborated.

(1) *MZ versus DZ (same environment)*

Most twin studies have made comparisons between MZ and DZ twins reared together. In addition to pointing to the relatively high correlations between MZ as compared to DZ twins, a number of investigators have utilized a formula (or derivatives) suggested in the Newman, Freeman, and Holzinger study (1937). This formula has been used to express the intrapair differences in terms of variance (sums of squared deviations from the mean for the pair divided by the number of pairs). One then calculates ratios of the variance observed in DZ twins (V_d) and MZ twins (V_m). From these are derived coefficients of heritability (H) from the formula:

$$H = (V_d - V_m)/V_d$$

When H equals zero the variations are considered to be purely environmental; when H equals 1 they are purely genetic, and intermediate values are claimed as estimates of the relative contributions of either heredity or environment. As already noted, the coefficients of heritability (H) turned out to be 65 to 80 percent; in other words, roughly two-thirds to four-fifths of the variance can be ascribed to heredity. This corresponds to the approximate correlational value of .90 of MZ twins frequently reported in the literature by proponents of the genetic position, and thus explains their claim that 81 percent of the variance, the correlational value squared, can be attributed to hereditary factors. However, this is only one side of the issue and, as will be suggested here, a biased and seemingly incorrect one. This becomes evident in formula 2.

(2) MZ (same environment) versus MZ (separated environment)

There is a second apparently disregarded and/or overlooked formula suggested by and utilized in the Newman, Freeman, and Holzinger study. Compared to the assumption of comparable environments for MZ and DZ twins in the estimations of H reported under (1) above, this method consists of an estimate of the importance of environment (E) in which are computed ratios of the variance of separated MZ twin pairs (V_{mz_s}) and unseparated MZ twin pairs (V_{mz_u}). Thus the formula:

$$E = (V_{mz_s} - V_{mz_u})/V_{mz_s}$$

Compared to the quoted estimate of heritability under (1) above, they reported as well the following for the estimates of the importance of environmental factors (all converted to percentage of the total variance): height 24, weight 87, head width 58, Binet IQ 59, Otis IQ 64, and Stanford Achievement Test 87. These, of course, are considerably higher estimates of the importance of environment than usually reported.

As further support for this position, the present author made similar comparisons from the data of Shields (1962). Ten separated MZ twins were matched for sex and age at separation and compared with ten MZ twins living together. On the Dominoes test the mean difference was 3.1 (greater for the separated twins) and the estimate of environment was 72 percent. For the Mill-Hill vocabulary the mean difference was 1.8 points and the estimate of environment was 62 percent. These values thus compare favorably with those reported by Newman, Freeman, and Holzinger.

Furthermore, and to add to the possibility that these estimates of the importance

of environment might be even greater, separated identical twins are not randomly placed in diverse environments. The selection of foster homes by agencies gives preferences to families who have sufficient financial resources to adequately care for the child and who show signs of intellectual and emotional understanding of the child's needs and the problems of adoption. Consequently, separated MZ twins placed for adoption through a professional agency are placed in selective and relatively homogeneous home environments as compared to the diversity that would result from random placement. Thus, the interpair MZ differences of 8.2 (and, inversely, the interpair correlation of .67) reported in the Newman, Freeman, and Holzinger study for separated MZ twins, could be even greater and thereby further enlarge the estimations of the importance of environment.

Moreover, and in addition to the selective tendencies of adoption agencies, there is an even more potent bias in the placement of separated MZ twins which would artificially enhance the genetic estimates and lower the environmental ones. Although most studies do not indicate very clearly (if at all) the nature of the families with whom the separated MZ twins are placed, the detailed work of Shields (1962) does provide adequate data for the point to be emphasized here, namely, that placement of MZ twins with relatives of the family also provides a more homogeneous environment than that found in the general population. Thus, as a specific example, if the mother of MZ twins retains one twin and the second is placed with a maternal aunt, certainly a greater similarity of child-rearing practices would exist between the two than if one twin had been randomly placed in another home. Two-thirds of the separated MZ pairs in the Shields study were placed with relatives of the family in which the MZ twin pairs were born. Again, the interpair MZ difference (and, inversely, the interpair correlation .77, Shields) would be even greater; and, the alternative means of deriving results from the formula of E greater than the percentages of 59 and 63 (Newman, Freeman, and Holzinger, 1937) and the 62 and 72 percent estimates from the data of Shields (1962).

(3) *MZ and DZ (separate environment)*

A third formula, and a means of limiting the confounding of genetic and environmental factors in the estimation of H, and also a method recommended in the Newman, Freeman, and Holzinger study (1937) but apparently disregarded by subsequent investigators, is to utilize the formula in (1) above, but with both separated MZ and separated DZ twin pairs. Assuming the control of such factors as the correct determination of zygosity, age of separation, same-sexed DZ pairs, and the random placement of the individual twins in diverse environments at an early age, to

name but a few relevant methodological considerations, a determination of the contribution of H may be more accurately assessed. Unfortunately the data available on separated DZ twin pairs is practically nonexistent. Jensen (1968) reports that the median value for separated DZ twins is approximately .42 (in his graphic illustration, page 50). The work of Erlenmeyer-Kimling and Jarvik (1963) to which Jensen refers *does not* report any such data. The only study of separated DZ twins known to this author is that of Shields (1962). Of 11 separated DZ twin pairs, Shields reports scores on only 4 pairs, hardly adequate data for useful analysis.

It is rather surprising, in view of the long history of the problem, the numerous publications generated, and the specific recommendation of the classic work of Newman, Freeman, and Holzinger (1937), that separated DZ and MZ twins have never been compared. Supposedly, separated DZ cases should be more readly available than MZ ones.

(4) *Alternative Analysis (separated MZ versus separated siblings)*

Whereas data obtained from separated DZ twins is either presently unavailable or has escaped my attention, a substitute may be considered for a reasonable analysis of the problem proposed by Newman, Freeman, and Holzinger (1937). Instead of comparing separated MZ and separated DZ twins as a means of estimating the H coefficient, and thereby controlling for confounding environmental variables, a similar although less satisfactory analysis may be computed between separated MZ twins and separated siblings. Of course, separated siblings do not provide for the most adequate analysis from the environmentalist's position since separated siblings of the same family are usually of different ages at separation, possibly of different sex, exposed initially to different family economic circumstances, and variable child-rearing practices during their early development, to name a few of the factors that would be less influential if separated DZ twins were available for analysis instead. However, even this method provides findings which are contrary to those commonly quoted in support of the hereditarian position.

As in formula (1), the assumption—although not altogether tenable as indicated—is that MZ twins and siblings would have an equally similar environment if separated in early life and that differences between the separated pairs and the derived variances could be ascribed to H. Thus in accordance with formula (1), the ratios of the variance observed in separated siblings (V_{sibs}) and separated MZ twins (V_m) can be used to estimate the coefficient of heritability (H) from the formula:

$$H = (V_{sibs} - V_m)/V_{sibs}$$

While these values are not readily obtained from most published studies, they may be derived from the following and essentially equivalent formula using the Fisher intraclass correlation coefficient (r):[2]

$$H = (r_m - r_{sibs})/(1 - r_{sibs})$$

Taking the median value of .47 for the 33 studies of separated siblings reported by Jensen (1969), one may arrive at H for each of the major studies involving separated MZ twins. The intelligence test correlations and estimates of H are presented in Table 1.

TABLE 1

Estimation of H with Separated MZ Twins and Separated Siblings.

Author	*Type Test*	*Correlation of MZ Pairs*	*Estimates of H (percent Variance)*
Newman, Freeman, and Holzinger (1937)	Binet IQ	.67	37.74
	Otis	.72	47.17
Shields (1962)	Mill-Hill Vocab. and Dominoes	.77	56.6
Burt (1966)	Binet, Pinter-Paterson, and Teacher's Report	.86	73.6
	Group Test (?)	.77	56.6
Jensen (1969)	Median Value	.75	52.8

As can be noted from a perusal of this table, the estimates of "intelligence" attributable to H range from 38 to 74 percent, and the median value reported by Jensen equals 52.8 percent. Similar estimates of H are presented in Table 2 for achievement test scores of separated siblings and separated MZ twins (derived from tables of Burt, 1966). The estimates of H are generally quite small, and the cor-

[2] This formula is based on Jensen's usage, where, unlike the Pearson product-movement correlation, the intraclass correlation is used as a direct estimate of the proportion of the variance accounted for by H; and, consequently, the values need not first be squared as in the formula:

$$H = (r^2_m - r^2_{sibs})/(1 - r^2_{sibs}).$$

The latter formula would result in lower H estimates than provided in the tables.

The intraclass correlation coefficients used to estimate the various proportions of variance differ from the well known product moment correlation coefficients, although Jensen never makes this distinction directly. The intraclass correlation was developed by Fisher and clearly described in his classic, R. A. Fisher, *The Design of Experiments* (London: Oliver-Boyd, 1935).

TABLE 2

Estimates of H on Achievement Tests with Separated MZ Twins and Separated Siblings.

Author	Type Test	Correlations MZ	Correlations Siblings	Estimates of H (percent of Variance)
Burt (1966)	Spelling	.597	.49	20.98
	Arithmetic	.705	.56	32.95
	General Attainment	.623	.526	20.46
Newman, Freeman, and Holzinger (1937)	Stanford Achievement	.507	—	—

relation between separated siblings on the general attainment score (Burt) is greater than that between separated MZ twins in the Newman, Freeman, and Holzinger study.

Concluding Remarks

Thus, when the effects of E variables are more adequately controlled, estimates of the importance of heritability in the determination of individual differences in intelligence and academic success are considerably less than suggested by Jensen (1969). Moreover, with separated DZ pairs instead of siblings, the estimates of H could be even less.

Although these findings are based upon formulas suggested in the Newman, Freeman, and Holzinger study, and especially the first utilized by later investigators as well (Eysenck and Prell, 1951; and Osborne, Gregor, and Miele, 1967), one potential criticism deserves elaboration. The claim may be forwarded that the correlation of .47 (Jensen, 1969) between separated siblings reflects "largely" the importance of genetic factors and thus confounds the estimates of H presented here. However, the reported findings in the literature are inconsistent on this point. For example, Freeman, Holzinger, and Mitchell (1928) found that siblings correlated .25 after 7 years of separation and, when foster homes of different grades were considered, the correlation between separated siblings was only .19. The immediate and customary reply to these latter findings is that potentially this relationship is still closer to .50 and the discrepant findings on siblings reared apart are due to differences in the opportunity to learn and/or to a lack of exposure to equally stimulating environments. But this is to beg the question. It is to assume the very point (that potentially

siblings correlate .50) raised in the question, that is, do genetic factors account for the similarities between related individuals?

Another means of sidestepping findings contrary to the genetic position is to argue that an intelligence test is not an adequate measure of intelligence or of the hereditary predisposition, potential intelligence. But then "potential intelligence" requires definition and measurable operations other than the tests commonly used. Moreover, to argue in this manner is to be inconsistent both with regard to the operational definition of "intelligence" usually ascribed to and the interpretations usually made from such studies.

In any case, and regardless of the large number of studies which support the genetic position, to the extent that they are subject to the methodological difficulties suggested here, these studies offer limited support to Jensen's claims. In fact, the findings reported here in which these methodological difficulties have been more adequately controlled are contrary to the claim that individual differences in intelligence and academic success can largely be attributed to hereditary factors.

The conclusion favoring the importance of heredity in the determination of intelligence has been with us most perceptively since Galton (1883) stated from his family study of eminence "that the instincts and faculties of different men differ almost as profoundly as animals in different cages of the zoological gardens." The confounding of environmental and hereditary factors has been with us equally long and it would seem that Galton's apprehension was justifiable:

"My fear is, that my evidence may seem to prove too much, and be discredited on that account, as it appears contrary to all experience that nurture should go for so little." (p. 241).

References

Burt, C. The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart. *Brit. J. Psychol.*, 1966, 57, 137-153.

Erlenmeyer-Kimling, L., and Jarvik, L. F. Genetics and intelligence: a review. *Science*, 1963, 142, 1477-1479.

Eysenick, H. J., and Prell, D. B. The inheritance of neuroticism: an experimental study. *J. Ment. Sci.*, 1951, 97, 441-465.

Freeman, F. N., Holzinger, K. J., and Mitchell, B. C. The influence of environment on the intelligence, school achievement, and conduct of foster children. *Yearb. Nat. Soc. Educ.*, 1928, 27, 103-217.

Galton, F. *Inquiries into human faculty and its development*. London: Macmillan, 1883.

Hempel, Carl G. Fundamentals of concept formation in empirical science. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press, 1952.

Herrman, L., and Hogben, L. The intellectual resemblance of twins. *Proc. Roy. Soc. Edinburgh,* 1932-3, 53, 105-129.

Jensen, Arthur R. How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* 1969, 39, 1-123.

Newman, H. H., Freeman, F. N., and Holzinger, K. J. *Twins: A study of Heredity and Environment.* Chicago: University of Chicago Press, 1937.

Osborne, R. T., Gregor, A. James, and Miele, Frank. Heritability of numerical facility. *Perceptual & Motor Skills,* 1967, 24, 659-666.

Shields, J. *Monozygotic Twins (Brought up apart and brought up together).* London: Oxford University Press, 1962.

Wilson, P. T. A study of twins with special reference to heredity as a factor determining differences in environment. *Hum. Biol.,* 1934, 6, 324-354.

Correspondence: Political, Technical, and Theoretical Comments*

E. N. ANDERSON, JR.

The Social Factors Have Been Ignored

To the Editors:

Certain rather sensationalistic accounts in the popular press directed my attention to Arthur Jensen's article. The article will no doubt receive comments from persons more qualified than I in psychology, testing, and education. However, as an anthropologist, I can raise some points that may not be mentioned by other workers.

Jensen's work is based on two assumptions: 1) IQ tests are a valid measure of inherited intelligence; 2) blacks and whites represent biologically different races—each more or less homogeneous within itself—in the United States. The second is more obviously debatable, yet Jensen does not discuss it.

In any biological sense of the word "population," blacks and whites do not constitute separate populations in the United States. Insofar as they can be called "races," the term is being used either in a purely socially defined way, or to refer to certain superficial features (notably skin color, nose shape, hair) which may

* The following correspondence has been selected by the editors from responses received concerning Arthur R. Jensen's, "How Much Can We Boost IQ and Scholastic Achievement?" (HER, Winter, 1969).

Harvard Educational Review Vol. 39 No. 3 Summer 1969

or may not relate to other hereditary entities. A biological population, on the other hand, is defined by breeding: its members interbreed with each other more frequently than they breed outside the group, and there is some sort of boundary —usually geographic—separating them from the out-group. Therefore the members of a population tend to share genes with each other more than they do with outsiders. Essential to defining a population is some measure of who mates with whom. Discussion of hereditary statistical differences is meaningful only in the context of well-defined populations.

Blacks and whites do not represent different populations in the United States —nor do poor whites and rich whites—because they do not fulfill these conditions. The social labels are not based on allocation to a gene pool. Many individuals have been classed as "black" at some time in their lives and as "white" at some other time (as when moving from an area of light-skinned people to an area of darker-skinned ones). The frequency of "passing" is high; it has been calculated that most Americans with some African ancestry are defined as "white".[1] More to the point, a mating between a black of one area and a white of the same area is usually more probable than a mating between persons of the same "race" but of widely different geographic residence. (Claims have been made that blacks and whites do not often interbreed. This is clearly wrong. In part it may be based on some confusion between local marriage norms and actual behavior patterns.) Therefore, to the extent that IQ is inherited, it will be inherited within the New York population or the Central Los Angeles population or whatever the genetically defined pool may be—not within the black or white races as Jensen uses the terms. The fact that individuals in city X have a low IQ has very little relevance (if any) to individuals of the same "race" in city Y, *whether or not the IQ scores are due to heredity*, unless there is extremely frequent and regular gene flow between the cities. The policy implication is that if IQ is indeed shown to be primarily inherited then we must determine IQs city by city, area by area, population by population, and educate accordingly.

Some other things follow from Jensen's use of socially or culturally defined groups as pseudo-populations. Scientific measures of heritability cannot be meaningfully used, since they are developed for use on true populations. If a pseudo-population is defined by reference to cultural traits X, Y, and Z, then a measure of heritability will always turn up the fact that traits X, Y, and Z are inherited,

[1] Robert P. Stuckert, "Race Mixture: The African Ancestry of White Americans," in *Physical Anthropology and Archeology: Selected Readings*, ed. by Peter Hammond, (New York: Macmillan, 1964).

because parents teach children. Language and dialect, bicycle riding, drinking behavior, political party affiliation all correlate quite well between parents and children. Jensen's misuse of heritability measures would allow us to conclude that any and all of these are inherited. I suspect that identical twins raised apart speak the same language in almost all cases, since adoption agencies and other placing bodies very rarely place twins in two different linguistic groups. I suspect that there is a much better case for inheritance of language than for inheritance of IQ scores. (A test experiment is needed.) Yet no one, to my knowledge, believes languages are inherited as specific traits. Siblings in the same family also have a way of speaking the same language, down to peculiar turns of phrase not used outside the family.

This is relevant to Jensen's first assumption, namely, that IQ tests measure something called "intelligence" that is somehow inherited (without reference to biological populations). Let us gloss over, as Jensen does, the fact that Jensen begins his article by saying that intelligence is a unitary thing, *g*, and ends by saying that it is at least two things, "cognitive" and "associative" learning ability. The IQ tests measure *something*. What they measure is a point of question. By Jensen's own admission, they measure familiarity with the test and test situation; he caused a rise of 5 or 10 IQ points in an hour or so by allowing children to relax and play around between tests. IQ tests also measure fluency in the dialect the test is written in (or that the directions are spoken in, if the test is nonverbal). On arrival from Finland, as a child, the girl who is now my wife was given an IQ test in English. A few years later she was given another similar test. Her score on the second was some 100 points better than on the first. I have been present at classrooms in which IQ tests in English were administered to monolingual English speakers and nearly-monolingual Spanish speakers—and the results treated as comparable. The southern dialects spoken by blacks in most cities are so different from general American English that a black and a northern white have real trouble communicating. To my knowledge, little attempt has been made to test speakers of southern dialect in tests written in their own dialect. (Note that southern whites score low on IQ tests.)

IQ tests also measure motivation. Under what conditions of motivation were the IQ tests cited by Jensen and Shuey administered? The middle-class white child is taught that his whole life depends on his doing the best he can on standardized tests; he becomes highly motivated, and often test-wise as well. The lower-class child rarely is so convinced. He is also apt to be in poorer health and nutritional state. Thus one expects class correlation with IQ scores, especially when class and

skin color are both against the testee. One may ask whether the IQ tests cited were given under conditions controlled for health; if they were given by sympathetic blacks or by overtly hostile whites; if they were made to seem important to all students equally. These are legitimate questions that are not answered by Jensen. I have seen IQ tests administered in school contexts in which it was clear to me and to students I talked to that the examiner was not impartial. I propose experiments as follows: 1) extensive testing AFTER students are controlled for motivation and state of health; 2) testing students—black and white together—by an openly racist white, a neutral (stranger) white, and a black man (or perhaps a neutral black and an openly anti-white one); 3) testing under different situations: in a middle-class white preserve (public schools are usually believed to be so by black pupils), on neutral ground (if there is any left), and on a ghetto street. In the last case, non-ghetto residents might be at a considerable disadvantage.

Finally, I am struck by the very small size of the difference that is finally produced: "When gross socio-economic level is controlled, the average difference reduces to about 11 IQ points (Shuey, 1966, p. 519), which, it should be recalled, is about the same spread as the average difference between siblings in the same family" (Jensen, p. 81). It is also well within the percentage of IQ variation that Jensen allows the environment to control! It is also, on Jensen's own showing, only a very few points more than the difference between identical (monozygotic) twins reared apart! And this without even controlling for any but "gross" factors! I feel that Jensen has made an excellent case for the lack of any significant difference between blacks and whites in IQ.

I am also interested in the fact that Jensen is explicit about blacks, but says nothing about other races, except for a passing reference to Amerindians. Orientals—even with language-barrier problems—do amazingly well on IQ tests in California. One misses a discussion of this, to say nothing of the differences between blacks of different cities and of different parts of given cities. Given these differences, it seems quite possible to me that IQ, including whatever heredity it may have, varies between populations. But populations are not races, nor are they at all close to races as defined in Jensen's work. Furthermore, policy implications of any difference that may be found are much less than Jensen seems to think. An average difference is not an absolute difference. Individuals vary so widely in IQ that the vast overlap is more conspicuous than the slight average difference. What of the millions of individuals assigned to the overlap section on Jensen's own bell curves? Would he consign them to the ash-heap?

Correspondence
S. ANANDALAKSHMY
JANICE F. ADAMS

A final consideration regarding Jensen's article comes from social-anthropological concerns. I feel that anthropologists know enough of cultural operations to be able to predict the effect of Jensen's article. It was published in a prestigious journal, easily available to the public. Quite predictably, the press seized on the article, exaggerated the racialist claims and played them up out of context and out of proportion, and failed to pay attention to refutations. The next step will be political; certain groups will use these press stories to bolster their political and social messages. This will involve further exaggeration. The public, poorly trained in genetics, will be swayed; I believe that major segments of the population will be convinced that "science" has "proved" that blacks are innately inferior to whites. Meanwhile, the blacks will not allow Jensen's article to go unchallenged, and in the current inflammatory racial situation this could have fearful results. Berkeley militants have already taken up the cry "Fire Jensen!" The outcome will be an escalation in the current racial conflict. It seems to me that responsibility for this escalation will fall on the author and publishers of Jensen's article.

E. N. ANDERSON, JR.
University of California

An Alternative Heritability Estimate

To the Editors:

We would like to make the following points in response to Jensen's article.

The geneticist Cavalli-Sforza says that "heritability measurements are somewhat arbitrary and can be given in a number of different ways (p. 8)."[1] He suggests that alternative methods of computing the fraction of genetic variance for intelligence might be "less striking and might be between 40 and 60% *from the same data*" (emphasis ours).

Jensen's method of arriving at a heritability estimate is ambiguous even in the earlier article[2] that he cites. In his computations, he assumes a test reliability of .95, an estimate which is higher than most test users would accept. He presumably corrects for attenuation and for unreliability, but his calculations are not made

[1] Luigi L. Cavalli-Sforza. Problems and prospects of genetic analysis of intelligence at the intra- and interracial level. Paper read at the AERA, Los Angeles, February, 1969.
[2] Arthur R. Jensen. Estimation of the limits of heritability of trials by comparison of monozygotic and dizygotic twins. *Proc. Natl. Acad. Sciences*. 1967, 58, 1, 149-156.

available to the reader. If one were to apply the formula he gives for heritability to the data presented in Table 2 (p. 49) where $r = .87$ for MZ twins reared together and $r = .56$ for DZ twins same sex,[3] one gets

$$\frac{.87 - .56}{1.00 - .54} = .67$$

which is lower than .80 that Jensen derives.

The estimate of genetic influence also seems to be made on the assumption of uncorrelated environments. Given Jensen's own statement (p. 50) attributing the correlation of .24 in the intelligence scores of unrelated children reared to-

[3] There appears to be an anomaly in Table 2, where the data on siblings and MZ twins are presented under two sections each: reared together and reared apart; while the data on DZ twins is in two sections: same sex and different sex, with no reference to the rearing situation.

TABLE 2

Correlations for Intellectual Ability: Obtained and Theoretical Values

Correlations Between	*Number of Studies*	*Obtained Median r*[a]	*Theoretical Value*[1]	*Theoretical Value*[2]
Unrelated Persons				
Children reared apart	4	−.01	.00	.00
Foster parent and child	3	+.20	.00	.00
Children reared together	5	+.24	.00	.00
Collaterals				
Second Cousins	1	+.16	+ .14	+ .063
First Cousins	3	+.26	+ .18	+ .125
Uncle (or aunt) and nephew (or niece)	1	+.34	+ .31	+ .25
Siblings, reared apart	33	+.47	+ .52	+ .50
Siblings, reared together	36	+.55	+ .52	+ .50
Dizygotic twins, different sex	9	+.49	+ .50	+ .50
Dizygotic twins, same sex	11	+.56	+ .54	+ .50
Monozygotic twins, reared apart	4	+.75	+1.00	+1.00
Monozygotic twins, reared together	14	+.87	+1.00	+1.00
Direct Line				
Grandparent and grandchild	3	+.27	+ .31	+ .25
Parent (as adult) and child	13	+.50	+ .49	+ .50
Parent (as child) and child	1	+.56	+ .49	+ .50

[a] Correlations not corrected for attenuation (unreliability).
[1] Assuming assortative mating and partial dominance.
[2] Assuming random mating and only additive genes, i.e., the simplest possible polygenic model.

gether to "... the fact of selective placement by adoption agencies, that is, the attempt to match the child's intelligence with that of the adopting parents," it is surprising that he does not apply the same principle to MZ twins. One wonders if adoption agencies have different policies with regard to placement of twins and of unrelated children!

S. ANANDALAKSHMY
JANICE F. ADAMS
University of Wisconsin

Misunderstanding Compensatory Education

To the Editors:

In the first pages of his lengthy article Arthur R. Jensen concludes that compensatory education has been tried and apparently has failed; this conclusion plays an important part in his overall analysis of the problem of Negro underachievement in traditional public school settings. It provides him with a justification for arguing in favor of a much more non-environmental (genetic) explanation of the differences between black and white IQ and achievement than is currently acceptable to most educators and social scientists. This conclusion is also necessary in order to help validate his argument that *environmental* conditions are *not* at the center of the differences found between average achievement rates of children from each race.

Jensen fails, however, to demonstrate the validity of this conclusion inasmuch as he reaches it by an analysis containing, in our estimation, at least three untenable assumptions: (1) that most of the educational programs up to now offered as "compensatory" could in fact *compensate* for the social inequities causing the underachievement; (2) that these "compensatory programs," putting aside the issue of their potential redemptive powers for the moment, have indeed been *adequately implemented* in most of the schools in which they have been introduced and upon which assessments of their effects have been based; and (3) that the traditional structure of public schooling *is* effective for children exposed to it when they have the basic ability to learn.

Since we believe that a great deal of what follows in the article rests on these unwarranted postulates, we would like to discuss several aspects of them to explain our reservations.

First, based on the citations found in the article, Jensen relies on the Civil

Rights Commission's conclusion about the *effects* of compensatory education. Apparently his eagerness to accept the Commission's conclusion did not permit him to raise fundamental questions regarding *its* logic and "findings." A far more rigorous and dispassionate review of "compensatory education" programs done by Gordon and Wilkerson[1], which also found that compensatory education programs have had little effect, presented a very different interpretation, namely that, given the nature of the problem, these programs are not really compensatory.[2] In the words of the authors:

> Weaknesses and limitations in these programs have been stressed in order to call attention to the fact that we have not yet found answers to many of the pressing educational problems of the disadvantaged. To assume that we have the answer is to subject multitudes of children to less than optimal development. More seriously, to settle for the beginning effort now mounted is to lay the basis for the conclusion that children of low economic, ethnic, or social status cannot be educated to the same levels as other children in the society. This conclusion could be drawn because despite all of our current efforts tremendous gains are not yet being achieved in upgrading educational achievement in socially disadvantaged children. We are probably failing because we have not yet found the right answers to the problem. To act as if the answers were in is to insure against further progress. (pp. 178-79)

Most of the programs, if one gives them careful scrutiny, involve the specification *on paper* of various combinations of the following activities: spending money on new educational hardware, adding teachers with special training to conduct special classes such as remedial reading, developing after-school enrichment programs in the arts such as poetry and creative dancing, and inducing parents to attend their childrens' schools on a regular basis by developing a series of parental programs. If the basic causes of lower achievement were environmental, could such specific programs as these really overcome the obstacles? Although a discussion of the probable environmental causes is too complex to present in this brief statement, we maintain that one must question, as do Gordon and Wilkerson, whether this type of "compensatory" program could in fact ever overcome the effects of the historical oppression and continuing overwhelmingly negative *environmental* conditions to which these children are exposed. In short,

[1] E. W. Gordon and D. A. Wilkerson, *Compensatory Education for the Disadvantaged* (New York: College Entrance Examination Board, 1966).

[2] On page 108 Jensen does acknowledge this excellent review, quoting from it in a way that *he* maintains support his "genetic" perspective. He fails to note, however, the following, central part of the authors' critical appraisal of compensatory education programs, which runs counter to his argument.

whereas Jensen's argument assumes at face value the potential effectiveness of these programs, given the proposed causes upon which they are based, we think this assumption must be questioned.

Second, no matter how promising these compensatory education programs are on paper, the measurement of how well they operate in producing the desired rise in achievement rate is contingent upon how adequately they are implemented *at the school level.* We maintain, furthermore, that the proposal and acceptance of a program as it is spelled out on paper, the allocation of necessary funds, and the busy activities and pronouncements of school personnel must not be equated with adequate implementation of the desired changes embodied in the program. Thus, before one can argue that the program *itself* is no good, one must demonstrate that it has been adequately implemented.

Once more, careful scrutiny of program assessments strongly suggests that neither the *assessments* of the programs' effects, nor the Racial Isolation Commission's *review* of these assessments, nor Jensen's *analysis* of the Commission's review, nor even the Gordon and Wilkerson appraisal of compensatory education programs consider the possibility that in general these programs were ineffective because in general they were inadequately or inappropriately implemented. (None of the basic assessments measure with any accuracy the degree to which necessary implementation occurred.) However, Jensen's analysis assumes that adequate implementation did take place. Given the growing literature on the problems of successfully implementing organizational innovations, we think that Jensen's assumption is the epitome of naiveté in organizational analysis. Therefore, we question at this time the validity of his conclusion that these programs have not worked because of genetic conditions found within the children.

Third, in connection with the issue of program effectiveness, it is curious that Jensen does acknowledge in a footnote that evaluations suggest that Project Headstart did have noteworthy effects, but that these effects were lost after the children entered the traditional patterns of schooling in the first grade. Yet he fails to interpret this finding. He maintains that the traditional form of schooling is basically effective: "The interesting fact is that, despite all the criticisms that can easily be leveled at the educational system, the traditional forms of instruction have actually worked quite well for the majority of children." Jensen might have argued that the Headstart evaluation demonstrates something is inherently wrong with the underlying nature of these students, since even after they have been given an initial injection, they fail to achieve in the traditional school setting, which works for the majority of children.

What evidence is there that the *traditional form* of instruction acts as the prime determinant of how well children achieve in school? Indeed, available research tends to support an opposite interpretation, namely that without the presence of other conditions usually associated with family SES, the traditional form of instruction is basically ineffective in producing adequate cognitive achievement in children. In our estimation, this interpretation supports an environmental, not a genetic explanation for why disadvantaged children exposed to temporary compensatory education programs such as Headstart lose the initial gains in IQ and achievement when they are shuttled back to the traditional school setting. We believe that Jensen's conception that traditional schooling is effective does not permit him to reason along this line, one which we feel is more relevant, given the data available at this time.

In sum, we believe that Jensen's analysis falls short of its mark because it fails to deal with some very fundamental, perhaps sociological, issues. Are most compensatory programs really compensatory? Have "compensatory" education programs been implemented adequately and for a long enough time to permit one to look elsewhere for an explanation of why they are ineffective? Is it the traditional form that schooling takes which accounts for why the majority of children achieve "according to their capacities"?

Before Mr. Jensen suggests that we move on to a more genetic explanation of differential achievement and IQ, we believe that he must provide us with convincing answers to at least these prior questions.

MARILYN BERNSTEIN
Harvard University
JOSEPH GIACQUINTA
New York University

Black Student Union Statement

The following statement from the Black Student Union of the Harvard Graduate School of Education describes their reaction to the publication of the Jensen article:

To the Editors:

In publishing the article by Arthur Jensen, the Editorial Board of the *Harvard Educational Review* gave tacit support, whether intended or not, to the argument that Black Americans are genetically inferior. This question is, of course, far

more political than scientific. The *Harvard Educational Review* Board either failed to recognize the need for consulting Black students or faculty on this article, or deliberately excluded them.

In a recent edition of the *Review*, an article on computers by Professor Oettinger drew apparently solicited responses that were included in the same issue. Evidently the Editorial Board went to considerable greater effort to provide a fair presentation on this far less controversial issue than it did on the question of racial inferiority.

The B.S.U. seriously doubts that the question of Jewish inferiority, Irish inferiority, or any other racist-inspired argument would have been thrust so arrogantly into prominence by the current Editorial Board.

We strongly oppose the license which the Editorial Board has exhibited in this matter and we demand the right to respond at an appropriate time of our own choosing to the Jensen article and to institute safeguards within the existing structure which will prevent the future printing of racist literature that is directed at maligning Black people in this country and/or abroad under the aegis of the Harvard Graduate School of Education.

Who's Being "Reasonable" Now?

To the Editors:

It just isn't so—"Shock waves are rolling through the U.S. educational community over a frank and startling reappraisal of differences in classroom performance between Whites and Negroes" (taken from *U.S. News & World Report*, March 10th, 1969).

The great majority of white teachers already had preconceived notions about white genetic superiority.

There is nothing—absolutely nothing!—new or shocking about the genetic claims of embattled privilege; the claims are as old as men's inhumanity to men.

All through history, dominant and privileged groups—no matter what their color, race, religion, nationality, class, occupational level, or what have you—have claimed biological superiority; even the claim of moral superiority has genetic implications.

And all that stale nonsense about cognitive learning and abstract reasoning!

If Jensen were "intellectually honest" enough to research he could find hundreds of blacks throughout our 400 years of existence in America who were, or are, intellectually equal and even "superior" to many racists.

To name a few: Frederick Douglass, the great black abolitionist and leader; an "ex slave" with no formal education, his ability to conceptualize and reason abstractly would put most current white college graduates to shame. His 4th of July speech at Rochester, N.Y., in 1852, has the quality of intellectual genius.

And W. E. B. DuBois, who was the intellectual superior of almost every white man that America has produced; Carter G. Woodson, the black historian; Dr. Daniel Hale Williams, who performed the *first* heart operation; Dr. Howard Drew, the discoverer of blood plasma.

And James Baldwin, Malcolm X and two black psychiatrists, Price M. Cobbs and William H. Grier—the list continues.

As always, environment is talked about, but almost nothing is done about "equal opportunity"—which is the valid and realistic battle cry of all oppressed groups.

It might interest racists that at *no* time in history have oppressed minorities been worried about the false claims of superiority of embattled privilege; what really agitates them is that superior freedom, superior rights (both legal and otherwise) and "superior" opportunities are projected as genetic superiority.

What happens to their genes when they fall from power? The Anglo-Saxon British aristocracy, for instance?

Aristotle was saying, some 2200 years ago, that some are "born" masters and some slave. He was talking about white men! One of the causes of the French Revolution was the upper-class claim of genetic superiority (they didn't use the term, but that is what they meant); i.e., that white men at the top of the social ladder were born to rule and exploit white men at the bottom of the social ladder, and those in between.

For well over 100 years white capitalists claimed that they were "here" and white workers "there" because of genetic inferiority. They still do.

To move back in time, there would be no Marxism if white capitalists had not ruthlessly exploited those whom they felt were their biological inferiors: the unorganized white workers; or, at least that was the fraudulent justifying principle —prejudice—or secondary reaction and "afterthought"—for their actions.

Jensen says: "... because the possible importance of genetic factors ... has been greatly ignored, almost to the point of being a tabooed subject." Garbage!

He has become merely the latest high-priest of racism. What about Arthur Gobineau, Madison Grant, Houston Chamberlain, Lothard Stoddard H. W. Odum, C. C. Brigham, McDougall, Nathaniel Weyl [*The Negro in American Civilization,* 1960], and Hitler?

Instead of science, this "exaggerated" genetics is better described as a neurotic —or is it paranoid?—vain attempt to get blacks, and other deprived people, to quit fighting for equal opportunity and believe in the *Santa Claus* of automatic justice.

How does the capacity for abstract reasoning and conceptual learning solve the problem of white injustice, lynching, burning, murder, denial of equal opportunity, and the most barbaric intimidation the world has ever seen?

Blacks have enough reasoning power not to accept all the nonsense and jazz that racists tell them is the cure for the Jim Crow system; i.e., *patience, education, be nice, Christian charity,* wait for the "good will" of the master, etc.

Blacks "know" that no privileged group in all history ever gave up its superior advantages, gracefully. Reason, education, and the ability to deal with abstractions are meaningless (were the Jews under Hitler lacking in intellectual ability?) unless the oppressed are willing to struggle, fight, defend, and die!

If blacks get justice—and they will!—it is because reason enables them to see through all the deception, pretense, hypocrisy, and make-believe.

Jensen says: "Heredity ... plays some role in the heavy representation of Negroes in America's lower socio-economic groups."

That statement is unbelievable, when one considers the fact that absolutely nothing is said about the extreme deprivation that blacks have endured—300 years of the cruelest slavery known to mankind; 100 years of barbaric servitude, murder, lynching, burning, and intimidation, superimposed with an arrogant, savage con game. There was literally no intention of treating blacks as human beings; but, rather, they were to be exploited and kept in servitude by any and all means, legal and illegal.

The most hypocritical part of Jensen's statement is about individual qualities and merits: most whites regard blacks not as individuals, but as an undifferentiated mass, and he knows this.

Every sophisticated black knows that most whites have a vested interest in the Jim Crow system, and this need to defend privileges determines the motivation for biased research; it selects the methods of study; and it makes it predictable that the conclusions will be the ideology of racism.

As for being reasonable, it is, literally, impossible for most white men to be reasonable about racism: they are locked in the terrible contradiction of, first, deceiving themselves, and, then, futilely attempting to deceive blacks that they want them to have equal opportunity—while simultaneously handicapping them so that whites can be privileged.

Blacks will accept absolutely nothing—let's repeat: nothing!—as proof or evidence but complete equal opportunity.

What blacks need is not the white man's genes, but more and more of the spirit of rebellion against racism and injustice.

ROY L. BROWN
Chicago, Illinois

The Prenatal Environment Is More Than Genetic

To the Editors:

The purpose of this letter is to discuss a serious weakness in Jensen's arguments for a large genetic component in the variation of intelligence. It is expected that this point will be discussed by many critics of Jensen's article, but due to its importance it may deserve repetition. This letter also suggests other possible sources of environmental variation which might account for a great deal of IQ variation and which would be automatically classed as genetic variance by Jensen's "heritability" estimates.

On page 68 of Jensen's article he suggests that individual differences in prenatal environments "... account for a substantial proportion of the total environmental variance in IQ." However, Jensen does not appear to recognize that variance attributable to certain prenatal environmental factors would be classified as genetic variance in any "heritability" estimate such as the correlation between monozygotic twins raised apart. Such prenatal environmental factors would be those common to both members of a set of twins but varying between sets of twins. The "... individual differences in prenatal environment (that) could cause IQ differences in single born children ..." (p. 68) would for the most part be common for both twins of a set. Jensen appears to incorrectly include this prenatal environmental variance in the meager 20% he attributes to environment (on the basis of his heritability estimates). This falsely implies that *postnatal* environmental influences are even less important than the "heritability" estimates suggest.

Since prenatal environmental factors common to each twin completely escape estimates of environmental variance by "heritability" estimates, and, in fact, are classed as genetic factors, it is tempting to call on such prenatal factors to account for most, or even all, IQ variance. To do this one must account for the different IQ correlations for persons of different degrees of relatedness. This would require decreasing amounts of communality of prenatal environmental

factors for monozygotic twins, dizygotic twins, siblings, cousins, etc. Such an explanation of IQ correlations between various relatives cannot be completely rejected. The largest problem may be to account for much more similar prenatal environments for MZ than for DZ twins.

One class of prenatal influences not discussed in Jensen's article is the modifiability of the fetal central nervous system by stimuli of the prenatal environment. Such modifiability has been shown to be possible in studies where the fetus is conditioned to react to neural stimuli (e.g., Spelt, 1948). In addition, Salk (1962) has indicated that some form of auditory perceptual learning occurs prenatally since recordings of a human heart beat have a soothing effect on the neonate and also on older infants. This latter finding suggests the possibility that prenatal auditory imprinting may also occur to the mother's voice. If voices are soothing to infants, this might be an important factor in the development of infant speech and other verbal behavior. Difference in prenatal exposure to the human voice might thus produce differences in later speech development. Other prenatal stimuli might also be of importance in later development and such stimuli may vary in their amount from pregnancy to pregnancy.

If prenatal learning is important to future IQ, there is another mechanism that could account for differences in such learning. This would be the amount of arousal of the fetus as determined by the mother's arousal level. Some form of Yerkes-Dodson law may operate whereby effective prenatal learning is precluded by too low or too high arousal. A chemical arousal transmitter would appear to exist (Pitts, 1969) which could communicate the arousal of the mother to the child.

In conclusion, the basis for Jensen's claims for a large genetic component in IQ variance is unfounded. Furthermore, the prenatal environment which he does not interpret correctly could even account for all of the variance that is not the result of postnatal factors.

FREDERICK N. DYER
University of Alberta

References

Jensen, A. R. How Much Can We Boost I.Q. and Scholastic Achievement? *Harvard Educational Review*, 1969, **39**, 1-123.

Pitts, F. N., Jr. The biochemistry of anxiety. *Scientific American*, 1969, **220**, 69-75.

Salk, L. The importance of the heartbeat rhythm to human nature: Theoretical, clinical, and experimental observations. *Trans. New York Academy of Sciences*, 1962, **24**, 753-763.

Spelt, D. K. The conditioning of the human fetus in utero. *Journal of Experimental Psychology*, 1948, **38**, 338-346.

Is No-holds-barred Research Possible?

To the Editors:

Jensen's article is the most scholarly, comprehensive, and contemporary review on the genetics of intelligence that has ever been published. Because it is so well done and raises so many significant questions of both an empirical and practical nature, it is very likely to stand as a basic point of reference for many years to come.

Some reviewers, though, are almost certain to select one or more specific aspects of Jensen's lengthy thesis for special criticism and thereby appear to take exception to the whole argument. No doubt, too, some may even find a reporting error or two, as I have myself.[1] It indeed would be unfortunate, however, if readers were to make a final judgment about this work influenced solely by the kinds of uncertainties which may be found in virtually all forms of scientific inquiry, without weighing *all* the evidence.

Although social scientists admittedly tend to overlook the inheritance of intelligence, the idea itself certainly does not violate the senses of most people. One recent survey, for example, clearly demonstrates that the vast majority of adults, parents, school teachers, counselors, and even school children believe that intelligence tests measure, to a greater or lesser degree, what a person is born with, although at the same time they recognize that learned knowledge makes a difference, too.[2] Jensen is not saying anything essentially different. The contrary notion that "all children have similar potential at birth" is *not* widely shared, probably not even among psychologists and sociologists.

Jensen's discussion of race and intelligence obviously is a far more sensitive issue. Yet, he keeps the dialogue, as one should, on a scientifically "neutral" plane. He does *not* conclude, and this needs repeating, that the average difference between Negro and white distributions on intelligence is the result of heredity. Rather, he only *hypothesizes* that genetic factors may play a *part* in the determination of the difference, then presents some rather convincing evidence indicating that the hypothesis is at least "reasonable" and concludes that "we need more appropriate research for putting it to the test" and that "such definitive research is entirely possible *but has not yet been done*" (italics added).

[1] E.g., see the first sentence of the last paragraph on p. 76 where the correlations between SES and IQ (under two years of age) and between SES and IQ (beyond two years of age) are given as positive and negative, respectively. The data from which this conclusion is drawn, however, indicate just the opposite and it is reasonably clear that the author actually interpreted the data correctly but inadvertently reversed the labels.

[2] David A. Goslin, *Teachers and Testing* (New York: Russell Sage Foundation, 1967).

Nevertheless, despite these disclaimers, Jensen will be misread, misinterpreted, and misquoted (e.g., see *Newsweek,* March 31, 1969, p. 84). This is unfortunate for a variety of reasons but in part because it places a much heavier burden upon social and biological scientists who are just beginning to design collaborative studies which could provide answers to some of the important research issues the author has raised. If reviewers insist upon interpreting this paper as creating a "holy war between hereditarians and environmentalists" (as if such pure types actually exist), then we may wait still another generation for the kind of synthesis between the biological and social sciences that the answers to these issues undoubtedly require.[3]

I would like to comment on the implications of Jensen's point that " 'No holds barred' is the best formula for scientific inquiry." While I would like to believe that he is correct, I am not at all certain that he is. Yet, the "search for truth" probably is such a compelling force that the scientific community is not likely to stop prodding until it has more answers. What then is the danger in seeking the truth, particularly if inherited differences in cognitive learning actually are found between Negroes and whites and more specifically if these differences are very marked in the lower and upper ranges of intelligence? There is, I believe, a very real danger. Why?

Virtually all readers would agree with Jensen that persons should be treated on the basis of their individual capacities and performance and not on the basis of "irrelevant" criteria. Societies, however, simply never have been, are not now, and are not likely to be in the very near future, organized in just this manner. Although the tendency is much less pronounced when individuals have developed a close personal relationship with one another, in a great variety of situations people normally tend to respond to each other on the basis of "secondary cues"—a person's speech, his mannerisms, his dress, or his age—as well as, in many cases, the color of his skin. Such characteristics quickly convey, more or less faithfully, specific meanings to the "actors" and thereby tend to govern the outcome of their interaction.

Given these propositions, which may be found in any introductory textbook in sociology, if Negroes, on the average, are actually *proven* to be genetically "inferior" in intelligence to any marked degree, it is almost a certainty that this

[3] I am actually more disturbed about the potentially destructive responses that two other articles on the genetics of race and intelligence are likely to receive than I am about Jensen's article. Both are soon to be published in leading academic journals with which I am familiar—one written by a qualified geneticist, the other by a self-educated physicist, i.e., "self-educated" with respect to the issue.

"bit" of information will be added to the general catalogue of items of knowledge which each of us regularly stores away as useful guidelines in our daily conduct. Unhappily, no amount of extolling humanitarian and egalitarian virtues or referring to the "overlap" in IQ distributions could completely protect those "blacks" who do *not* fit the stereotype.[4]

Nor can sociologists probably find much comfort in the recommendations which Jensen and others have put forward that all we need to do is devise an educational system and occupational structure sufficiently diverse to provide for the development and utilization of all forms of human talent, plus a system of social rewards which does not discriminate one kind of talent from another. The assumptions upon which such false hopes are built are strikingly similar to those the Bolsheviks borrowed from the works of Karl Marx in planning a utopian, classless society. Social scientists and the Soviets alike have since discarded them.

First, it should be noted that it is the state of technology which largely determines the kinds of human talent that at any particular point in time a society finds useful to employ, and not the other way around. Second, in any free society which relies upon incentives rather than coercion to motivate and control human behavior, social rewards in the form of prestige, power, and wealth are going to be unevenly distributed. Consequently, those persons in positions that are more "functionally important" to a society and that require more in the way of one kind of talent than another are usually in greater demand. Being more highly valued commodities, they may even be called "superior." I frankly see no solution to this problem in the long run, except to consider more seriously than we have in the past some form of biological engineering *or* to prove Jensen's hypothesis wrong.[5]

BRUCE K. ECKLAND
University of North Carolina

[4] For an excellent discussion of the potential social consequences of a hypothetical situation in which science were to "prove" Negroes intellectually inferior, see Marvin Bressler, "Sociology, Biology, and Ideology," in *Genetics*, ed. by David C. Glass (New York: Rockefeller University Press and Russell Sage Foundation, 1968) pp. 178-210.

[5] E.g., see Frederick Osborn, *The Future of Human Heredity* (New York: Weybright and Talley, 1968).

In Praise of Jensen

To the Editors:

The publication of Arthur Jensen's splendid article—actually a book in all but name—is one of those signal events that are rare in any field of science: the appearance of a scholarly work that will for years, and possibly for decades, be regarded as the watershed that divides a period of misunderstanding, error, and myths from a new era when emergence of the true facts led to the formation of a solid theory upon which future scientific progress can be built. The Jensen article has already attracted more attention in other media and among the public than probably any other article in the *Harvard Educational Review*. It is certain to have a deep and lasting impact on several academic disciplines and on the thinking of the general public. Congratulations and thanks to the editors of the *Harvard Educational Review* for publishing it.

For some years it was almost unthinkable to mention differences among individuals or groups in intelligence or educational, occupational, economic, or other achievements without stressing in the same sentence, or in the next, the environmental factors which must have caused those differences to come into being. It was explicitly claimed or implicitly assumed that laws of heredity apply to all human characteristics and throughout living nature—but not to human intelligence. That assumption seems to answer a deep emotional need of some people to believe that all men are created equal in terms of intellectual capacity, that all children save for a few patent defectives are endowed with an intelligence that is close to the average and therefore can perform at that average or norm. Differences in achievements must then be attributed to environmental influences such as bad schools, poor homes, inadequate or defective societal mechanisms; in short, to anything but the inherent makeup of the individual.

Few would attribute all physical weakness, or inability to sculpture, paint, sing, or play basketball well exclusively to environment and inadequate training. It would strain credulity too much to assert that genotype has no bearing on performance in those skills. But when it comes to intelligence the true egalitarian must *a priori* assume the absence of the influence of heredity.

For many years such concepts not only dominated the speeches of ambitious politicians and newspaper front pages, they were also translated into multi-billion-dollar public programs. It is so much easier to comfort someone that he has been deprived by an unthinking or malevolent society than to tell him plainly that he just is not very bright.

When promises about the achievements of imaginative public programs (i.e., programs based on imagination rather than facts) went unredeemed they caused disappointment and frustration, then belligerence; and finally led to violence. Subsequent attempts to achieve the desired end by multiplying the amounts resemble nothing as much as the centuries-long quest of the alchemists, in the face of consistent failure, to convert base metals into gold.

Arthur Jensen's painstaking documentation and irrefutable logic have, hopefully, ended that pre-scientific period. Published comments on the article suggest that even those who are emotionally unable to accept Jensen's findings, must admit the steel trap nature of his facts and logic.

Some have attempted to soften the blow to their most cherished beliefs by tampering with the facts—and Joseph Alsop was leading among them. After paying tribute to Jensen's work, Alsop, in the second of two newspaper columns, undertook to deny the failure of compensatory education:

For example, Dr. Jensen includes the conventional educationist's sneer at the ill-success of New York City's Higher Horizons program in the high schools. He does not say, however, that Higher Horizons was the starveling, misbegotten offspring of the decidedly successful Demonstration School Project. And he does not point out, either, that one failed where the other had got results, because of a very drastic cut in the per pupil investment.

The *Demonstration Guidance Project* (that was its official name) in 1956 selected the *pupils with the greatest academic potential* at a Harlem junior high school (#43) and transferred more than half of them later to an academic high school (George Washington). Out of 717 pupils originally selected, 240 ultimately graduated.

The successor program, Higher Horizons (HH), took in all pupils (64,000) in 76 schools in poverty areas. It failed, not because of lack of funds but because of absence of any tangible results. After three years of operation its founder and coordinator proclaimed confidently:

It is not enough for us to raise the self-image, or to broaden cultural horizons, or to improve the school climate, or to make teachers happier. If within a reasonable period of time, the level of academic functioning has not been raised, then our efforts must be adjudged largely a failure.*

So they were. When subsequent investigation proved that there was no difference between the reading and arithmetic scores of pupils from comparable back-

* Jacob Landers, *Higher Horizons, a Progress Report*, N.Y.C. Board of Education, 1963, p. 9.

Correspondence
LESLIE A. HART

grounds who had attended HH schools and those who had not, the curtain fell on HH.

Nor is Mr. Alsop's attempt valid to discredit the evaluation report of the subsequent *More Effective Schools* program in New York City ("the outrageously slanted report condemning the More Effective Schools program comprising 21 New York primary and elementary schools" [sic!]).

That voluminous report was prepared at the behest of the New York City Board of Education by the Center for Urban Education, a New York research institution, federally financed under Title IV of ESEA for regional educational laboratories, and based on the research of a team of 38 New York educators and social scientists. The report contains all the test findings and statistical tables for anybody who cares to disprove its conclusions. Nobody has yet been able to do so.

Alsop claims that for cost reasons "systematic, radical school improvement" has never been tried in the U.S. and that "we shall never have ghetto schools that really educate until the federal government pays most of the bill for them." That seems to overlook that the federal government has, over the past four years, spent over $4 billion on Title I compensatory programs in over 60,000 projects designed and carried out by the initiative of local schools. Those projects cover a huge variety of goals, techniques, and approaches. Does none of them meet Mr. Alsop's concept of "systematic, radical school improvement"? What would he regard as such if doubling the amount per pupil *from federal money* (which presumably is far more educationally effective than plain old-fashioned state-local money) did not accomplish much if any progress?

The most frequently heard claim is that we are not starting early enough with compensatory education. We were told that we need to start at age 5 with kindergarten, then that we should intervene at age 4 with prekindergarten or headstart, then at age 2 and finally at birth. Considering all the factors, does it not seem likely that intervention at birth may come about nine months late?

ROGER A. FREEMAN
The Hoover Institution

This "Science" Has Nothing To Do With Schools

To the Editors:

Dr. Arthur R. Jensen's stimulating but unfortunate paper quite properly suggests that a study of genetic factors in relation to IQ may be of scientific interest.

But may I submit that it is most improper to raise the question in regard to *schools*. Insofar as a school or a teacher deals with any individual student or small group of students in the light of skin color or ethnic origin, racism and prejudice are introduced that we must not tolerate.

The author himself points out (page 78): "The variables of social class, race, and national origin are correlated so imperfectly ... that these background factors are irrelevant as the basis for dealing with individuals—as students ..."

One might wish that Dr. Jensen had at this point listened to himself and recast his article. Instead, he seems repeatedly to suggest that black students may need different treatment *because they are black*. Anyway you slice this, it is racism.

Quite the most amazing statement in the paper appears on page 7: "... the traditional forms of instruction have actually worked quite well for the majority of children." No support whatever is offered for this vague, sweeping claim. For any recent period particularly it is, I believe, quite absurd. Our schools have always depended on the whip, systematic fear, and despotic authority to operate at all. They have taught millions that learning is a dull, dirty business to be terminated at the first opportunity—as soon as one has an appropriate "union card" in the form of a diploma. Far from being concerned with *g* in any real sense, they are notoriously anti-intellectual, belittling, and suppressing intellectual activity as troublesome in the classroom, in favor of remembered right answers and approved algorithms. Dr. Jensen seems to regard time-serving, exam-passing, and compliant behavior as "doing well" and synonymous with "academic achievement."

The idea that most children do well in school, but that some (mostly black) do badly because there is something wrong with *them* has caused enough mischief—and Dr. Jensen's approach tends to add to it. The truth appears to be that the traditional school is, in John Gardner's phrase, "monumentally ineffective." By and large the schools discourage more learning and intellectual growth than they promote—with the effect that "achievement" neatly matches the child's resources *outside* of school. Those children who lack such resources do poorly, in proportion to their lack, because the school's "teaching" long has been a folklore-based ritual incapable of accomplishing anything on its own.

Dr. Jensen surely wins the prize for the most tenuous, dubious, and far-fetched argument yet, to prove that our antique, collapsing schools are right—and the children wrong. One shudders to think of the people he comforts.

LESLIE A. HART
New Rochelle, New York

Potential Is Not Measured by Performance

To the Editors:

A. R. Jensen in his lengthy discussion suggests that differences in IQs such as those between social classes, between Negroes and whites, and, incidentally, between Jews and non-Jews are accounted for mainly by genetic factors. Therefore, many such group differences in intelligence, Jensen suggests, cannot be affected by environmental changes.

Jensen explains that the only operational meaning there is for "intelligence" is, in effect, that which IQ tests measure (page 5 and following pages). He shows that IQ tests are originally based on what teachers expect from children in school; that IQ tests are "shaped by the educational traditions of Europe and North America."

Jensen then discusses "heritability"—the extent to which a phenotypic phenomenon (a property like height, a behavior like responding on an IQ test, and so on) is referrable to genetic factors. Jensen makes it clear that high "heritability" does not necessarily imply immutability: when there are changes in the environment, the extent of genetic influence on phenotypic phenomena may change.

Referring to analysis of variance, to correlations of IQ scores in people of various degrees of relatedness, to studies like the breeding of maze-bright and maze-dull rats, and others, Jensen shows that genetic factors are at work in behavior represented by the IQ, and that "heritability" is high.

Jensen concludes that, in the main, attempts to equalize IQs are doomed to failure; that, for example, Negroes on the whole inherently possess less of a certain kind of intelligence (viz., abstract) than whites.

The entire problem is, of course, largely academic—in the true sense: as Jensen shows (page 78), whether group differences are inherited totally, somewhat, or not at all, is irrelevant for social action and policy as long as there is overlap, since in practice we always deal with individuals.

Nevertheless, it is of course important—especially at this time—whether scientists give the public the impression, or whether they operate on the proposition, that scientific analysis shows that some socially identified groups are inherently, therefore unalterably, more stupid than others.

We contend that Jensen has not made such a case; that, indeed, his own definitions and considerations fail to support such a conclusion.

We will not deal with details of Jensen's discussion and treatment of evidence:

his slipping in value judgments of what is "important" or "large"; his starting by declaring compensatory education a failure, ending by citing evidence of its success; his using the maze-bright and maze-dull rat breeding studies as support for his strong-heredity hypothesis (page 31), yet illustrating by the follow-up (page 40) the startlingly strong effects of environment here, which well-nigh wiped out the inherent differences; his using a difference between correlation coefficients (non-linear, of course) to arrive at a proportion of variance (page 50); his reporting a higher "heritability" for IQ than for scholastic achievement, the former primarily predicting the latter; his strictures on "unbridled environmentalists" (page 29), whom no serious scholar is likely to take seriously; his contradictory discussion of whether IQs and intelligence can or cannot be raised (page 100), and what this means; his citing a paper in support of his thesis, omitting quotes from the same paper flatly (and pungently) contradicting Jensen; and others.

We shall concentrate on the concepts "intelligence" and "heredity"; confusion, error, and needless debate seem to arise largely when these concepts are misused and misunderstood (as by Jensen, despite his originally careful definitions).

Difficulties with "intelligence" often arise from its different meanings, which often are not kept distinct. First, it can refer to some presumably fixed capacity, ability, potential. It refers to what one might do, at most and at best, on tasks defined in the culture as requiring intelligence. This potential must, in the last analysis, be inherited. We refer to this when we define intelligence, say, as "the capacity to learn." (A close analogy is our saying that generally a six-footer probably can, is potentially able to, jump higher than a four-foot-tall man—regardless of how high they now do jump, or, indeed, whether they now jump at all.) We might call this "intelligence—potential ability."

Second, we speak of "intelligence" in terms of people's actual performance on tasks defined as intelligent: Einstein, Jensen, the valedictorian are "intelligent." We might call this "intelligence—performance."

Third, we use "intelligence" as that which is measured by IQ tests, which are correlated with and estimates of "intelligence—performance." We might call this "intelligence—performance estimate."

The essential point is that IQs ("intelligence—performance estimate") derive from behavior defined as intelligent. They are, and cannot be anything but, performance measures. As such, they have, of themselves, nothing whatever to do with "intelligence—potential ability"; they do not measure any kind of possible maximum capacity—no performance ever does (it indicates *minimum* available

potential). If you see me jump four feet high, you know I can do at least that; from this performance you know nothing about how high I might jump—what my inherent potential is. The same is true for groups, and is not affected by any estimates of how much genetic factors have contributed to this present performance.

IQ tests measure performance. Any conclusion as to potential ability or capacity is always an inference, which cannot meaningfully be made from the test score alone; and it is highly tentative. If I am a healthy, seven-foot-tall, twenty-year-old Watusi jumping four feet high, you will suspect that I am performing considerably below my potential (jumping) ability; if I am a sixty-year-old Pygmy, you will suspect that I am performing close to my maximum capacity, that I probably cannot jump much higher under any circumstances.

As mentioned, Jensen defines "intelligence" as "intelligence—performance estimate." But this could not give any information about potential ability. So, explicitly and implicitly, he keeps falling into the basic error of feeling that the IQ is, after all, a measure of "intelligence—potential ability."

If only we had never called these things "*intelligence* tests"! Failing this, if we could only curb our emotional involvements and stick with Jensen's original explanations and definitions of IQ and intelligence!

The IQ is an estimate of performance on tasks defined as requiring intelligence. The IQ provides an estimate of minimum (demonstrated, expressed) capacity or ability. The IQ provides no information, by itself, as to what an individual or group *might* do, what the potential ability is.

Incidentally, any potential capacity—to perform intelligently, to grow, to sing, to jump, and so on—is a quite highly abstract construct. It is not like the "capacity" of a quart jar for holding water; there, "capacity" is a quite low level abstraction. IQs are not analogous to cubic inches!

Confusion also attends the concept "heredity." It is more general than "heritability," which, as mentioned, is the extent to which a property like height or blood type or a behavior like learning mazes or responding on IQ tests—phenotypic phenomena—is referable to hereditary factors, to the genotype. Jensen makes clear that "heritability" does not imply immutability.

Yet, it is on the basis of such "heritability" that Jensen finally implies that IQ differences between groups indicate inherent and largely immutable differences in certain kinds of ability.

The difficulty arises because Jensen seems to think of heredity in the old, naive sense of the genetic transmission of specific properties: as if we directly or indirectly inherited, say, blood type, stature, and behaviors.

If we did, it would make sense to inquire as to the extent to which some phenotypic phenomenon is inherited, and to what extent it is due to environment. Any given phenotypic phenomenon could then theoretically be due entirely to heredity, entirely to environment, or to some specifiable mixture of the two. This way, we think of the phenotypic phenomenon as something like a martini, made up of gin and vermouth in varying proportions.

However, observed events make it impossible to think meaningfully of heredity and environment in this way. For example, presumably girls inherit something to do with menarche; but the age at menarche has decreased in a straight line for over a century (Tanner, J.M., *Growth at Adolescence,* Blackwell Sci. Publ., Oxford, 1961). Thus we "knew" a hundred years ago that the average girl inherited menarche at about sixteen; only, now we "know" that environment accounted for that—actually the genes seem to say "thirteen." And what the inherited possibility, the actual limit, is, we have no idea of. (This "secular trend" may conceivably be due to a sudden, simultaneous mutation all over the shop, though no one seems to have suggested this seriously; even so, there would presumably have been something in the environment "causing" such mutations.)

If anything may be taken as inherited, surely it is species-specific behavior—built-in, programmed, reflex, instinctive. What then of a breed of fighting mice, which, when reared by non-fighting rats from shortly after birth to weaning, do not fight? (Denenberg, V. H., Hudgens, G. A., and Zarrow, M. Z., "Mice reared with rats: effects of mothers on adult behavior patterns," *Psychol. Reports,* 1966, 18, 451-6.) We "know" that these rat-reared mice have inherited the capacity to fight—we have seen it in all mice of this breed. What do we know of inherited capacity to fight of animals we have never seen fighting? (Or of the capacity of lower-class children to abstract, relatively few of whom we have seen doing so?)

Jensen himself cites the finding that, after maze-bright and maze-dull rats have been *bred,* a very restricted environment results in both equally performing very poorly; and in a very enriched environment, the bright rats perform still better than in the ordinary environment, the dull ones much better, so that here differences nearly disappear (Cooper, R., and Zubeck, J., "Effects of enriched and restricted early environment on the learning ability of bright and dull rats," *Canad. J. Psychol.,* 1958, 12, 159-64). But this means that the genetic effect here is visible only in what happened to be the average environment in our laboratories. Had the enriched environment been the norm, we might never have discovered that maze-brightness and -dullness could be bred!

In addition, cell specialization right at the beginning is seen to be a function

not only of heredity, but of the influence of cells on each other, of the intercellular environment.

In view of these kinds of phenomena, it seems ineluctable that we give up the notion of the inheritance of phenotypic phenomena. We inherit possibilities, tendencies, limits—and these are inferred, though necessarily: even without Jensen's impressive citations of geneticists and others, it is clear from elementary considerations that any phenotypic phenomenon must be influenced, directly or indirectly, by genetic factors.

Any phenotypic phenomenon occurs when an organism's inherent potential meets an environment facilitating its expression. Obviously no environment can bring out anything if the potential for it does not exist—presumably we are all inherently incapable of flapping our arms so that we fly. But rarely is this potential directly or even indirectly observable in itself. Of course, it is clear that, for the organism to develop at all, some potentials will practically always have to be available, and an environment facilitating their expression will have to be present: thus, it is not entirely "wrong" to speak of the inheritance of blood types; of noses, arms, and legs growing in their accustomed places, *et cetera* (but recall the Thalidomide tragedies, where environment was changed so that the inherent potential to grow arms and legs could not be expressed).

The essential point here is that, in the area of heredity and environment, we are dealing with possibilities and their chances of realization. Thus, as Dobzhansky also says in the paper quoted by Jensen, no phenotypic phenomenon is uniquely and definitely "determined" by heredity; any phenotypic phenomenon is also genetically "conditioned"—it implies some inherited potential (Dobzhansky, T., "Genetic differences between people cannot be ignored," *Scient. Res.* July 22, 1968, p. 33).

In other words, phenotypic phenomena are not the "product of heredity and environment" in the sense of that martini "made up" of gin and vermouth; they are more like the music I hear coming from record and phonograph (the record may be as hi-fi as can be, but if the record player cannot bring it out, the result is poor, and *vice versa;* but except for some obvious scratch on the record, we have no way of telling whether shortcomings are a function of the record or the player—and some players can bring out tones and suppress noise to make fine music that another phonograph cannot produce with the same record).

This kind of what may possibly be called "interactionism" is—also on the basis of Jensen's own more careful considerations—a *sine qua non* for scientific understanding. The old nature-nurture controversies really are naive and out-

dated: there simply is no visible nature without nurture, no nurture without nature. And invisible nature, such as some absolute, reified inherent capacity, is not scientifically useful.

Rarely can we even indirectly measure inherited possibilities, especially intellectual, and never by a performance measure like the IQ.

All this does not of course remove the possibility that, for instance, Jews may have inherited tendencies making for Nobel Prize winning more than non-Jews. However, the fact that they do win more, or that some performance correlating with Nobel Prize winning may show high "heritability," does not "prove" this.

Jensen is, however, right in his call for educational opportunities of all kinds for all people. For IQ differences between groups, where there is considerable heritability, represent a challenge to create an environment where any socially or personally valuable potential that may have been inherited can be expressed.

WALTER HARTMANN
Purdue University

Intelligence Is More Than Measurement

To the Editors:

A. R. Jensen makes a valuable contribution by systematically coordinating a great many studies bearing on his chosen topic. It is only hoped that his conceptualizations of the data do not represent the extreme point of a swinging pendulum. Unusual interpretations are often signs of inadequate evidence which is then supplemented by wishfulness and self-projection. In behavioral research, the reliability of cumulated data such as are presented by Jensen is usually high, but the validity of causal inferences is nearly always subject to serious doubt.

Jensen's article is studded with a number of perceptive observations. These alone make the publication worthwhile and suitable for a stimulating exchange of ideas. Conceiving of intelligence as a "transfer" across behavioral functions is perhaps the boldest and most exciting idea from a clinical standpoint. Yet, Jensen fails to pursue this creative point. Instead, he reverts to the comfortable but discredited definition of intelligence as "the capacity to reason." Reasoning has been shown by Thurstone and many others to be a group factor independent of the *g*. Therefore, it cannot serve as a central vehicle for *g*. The existence of pure *g* tests, such as the Raven Progressive Matrices and others, is probably as unreal as is the

existence of culture-free tests. They could only exist in a social vacuum or on a planet inhabited by one-dimensional robots interacting with an unvarying environment. Fair measures of individual achievement on this earth may be possible at long range, but they will not be as simple as the search for panaceas in the form of pure *g* and "culture-fair" tests.

In a strained attempt to be consistent, Jensen creates an *ad hoc* hierarchy of associative (lower) and conceptual (higher) thought processes. His conclusion that the higher conceptual processes are measured by the IQ would be very difficult to demonstrate. Most IQ scales include a balanced number of associative and conceptual tests. A glance at the test manuals (Wechsler, etc.) and at various research papers reveals that the associative tests are more highly correlated with the IQ than the conceptual ones, except where the test design is an inferior one as in the digit span test.

Jensen's question directly concerned with the magnitude of the differential contributions of heredity and environment to intelligence is well taken. But there is another imperative question that must be asked and perhaps even answered before our concern with "how much heredity and how much environment" crystallizes into a research project. The question relates to how much the IQ and intelligence contribute to the totality of human behavior and achievement. It has been maintained by this writer for a number of years that the *g*, defined as the level of optimum personality integration, determines not more than 15 percent of the variance of the total life adjustment of any one individual. The remaining variance of 85 percent (including errors of measurement) is controlled by nonintellectual factors such as sensory acuities, linguistics, attitudes, self-concepts, fears, motives, moods, muscles, and yes, all the cognitive functions of concept formation, judgment, and reasoning. Should these group factors be modifiable by special stimulation and environmental reenforcement, then the heritability of intelligence is after all of little consequence to the individual and to the society in which he lives. The large number of failures with high IQs and the equally large number of successes with low IQs is fair evidence of the relatively small part which intelligence plays in overall learning. Jensen is of course aware of the possibility that intelligence is only a part of the total behavior complex. Still, he discusses IQs from different tests and different research projects as if they were absolutes and thereby imparts the erroneous impression that intelligence is equal to total personal accomplishment. In reality, neither Jensen nor those whom he quotes have measured the 85 percent variance hidden in the tests and not accounted for by the IQ.

Experienced clinicians know through retests that solid and permanent boosts in IQ are very gradual and take many years to materialize. One should not expect drastic changes in IQ as a result of intensified teaching that lasts 3 months or even 3 years. The validity of rapid and sustained rises of IQs should be viewed with great circumspection, as they are often due to poor tests or poor testing conditions either before or after the specialized treatment. Besides, the effecting of a rise in IQ by remedial education is a wholly misplaced and unimportant objective.

If a child's motivation for effort has improved or if a person's self-concept and reality contact have changed for the better, the attendant results may not be measurable by IQs and yet be decisive as to whether the individual's future effort will be positive or negative. Even those who may have been permanently damaged by early malnutrition, sensory deprivation, and gross neglect may benefit from compensatory training in the nonintellectual phases of behavior.

Jensen's curves and graphs illustrating the varying rates of development of the cognitive processes are partly confirmed by clinical retests. However, both associative and conceptual thinking are subject to negative and positive accelerations at different ages, in different groups, and for different reasons. The negative acceleration of children with speech and word recognition disabilities (associative processes) has been known for many years. The conceptual processes of these children are usually intact. They create amazement in teachers who observe the children's fine conceptual thinking in the presence of serious defects in the reconstructive skills. What is not so well known is that a prolonged period of positive acceleration in associative learning occurs in the same individuals between ages 16 and 30. Jensen's negative acceleration in conceptual thinking between ages 4 and 8 is repeated even more conspicuously in the years before puberty regardless of socio-economic status. It occurs in children with superior associative thinking and high IQs. Table 19 on page 115 is not to be accepted as universally valid. Indeed, it looks like the result of an experimental or statistical artifact based on biased tests, deviant population samplings, procedural idiosyncrasies, and inadequate definition of concepts.

In summary, intelligence should not be misidentified with the total adjustment of people, scholastically or vocationally. There are over-riding extra-intellectual forces which may cause competency or incompetency at any IQ level. The idea that intelligence is mostly a genetic trait (still to be confirmed) should not deter this country from initiating remedial projects intended to improve the quality and quantity of human self-realization. If only 12 percent (80% of 15) of total

behavior is inherited, it would be inhuman to deprive large segments of the population of a chance to learn and to contribute to the creative pool of the community. There is still much to be learned about learning, but this important task may be frustrated unless we place this IQ in proper perspective with regard to overall achievement. Studies like Jensen's tend, by their selective and exaggerated emphasis on minor issues, to narrow public horizons and to foster social inertia or what has aptly been named "education for retardation."

JOSEPH F. JASTAK
Wilmington, Delaware

References

Thurstone, L. L. Primary Mental Abilities, *Psychometr. Monogr.*, 1938, No. 1.

Thurstone, L. L. & Thurstone, T. G. *SRA Primary Mental Abilities.* Technical supplement, Chicago: Science Research, 1954.

Jastak, J. F., MacPhee, H. M., Whiteman, M. *Mental Retardation. Its nature and incidence.* Newark, Del.: Univ. of Del. Press, 1963.

Jastak, J. F., Bijou, S., Jastak, S. The Wide Range Achievement Test. Wilmington, Del.: Guidance Associates, 1965.

Jastak, J. F. Intelligence tests and personality structure. In *Psychopathology of Mental Development.* New York: Grune & Stratton, Inc., 1967.

Racial Alienation and Intelligence *

To the Editors:

Arthur R. Jensen's article is a thoughtful review that should be read and discussed by a far larger audience than is likely to see it in the Winter issue of the *Harvard Educational Review.* It will be much talked about, but unfortunately only secondhand in response to several popular commentaries that have emphasized a few controversial (and I would say incautious) remarks at the expense of a great deal of Dr. Jensen's wisdom and scholarly reserve.

The meat of his discussion concerns the effort to bridge the IQ gap between the white and Negro communities in the United States. There can be no evasion of the raw statistics, which indicate, among other things, an average reading retardation of one to three years. The question is whether we can design educational programs to erase the painful statistics.

Dr. Jensen is careful to insist that we focus on individual capability: genius is neither lacking among Negroes nor universal among whites. He does point out, wearily, that we cannot overlook the social demand for programs that concentrate on compensation for group handicaps.

His most provocative statement is his first sentence: "Compensatory education has been tried and it apparently has failed." Unfortunately such a remark may deter many proponents of the principle of compensatory education from reading the substance of his criticism. There is little doubt that many programs could not begin to meet the unrealistic expectations of their enthusiasts. In this sense, we could argue that every educational program has failed, and note that many brilliant men have achieved their successful place in life in spite of wholly inappropriate educational regimes. Many critics believe that compensatory education has hardly ever been tried, and within our present social framework it may be impossible to implement with the rigor needed to achieve prompt returns. Compensatory education programs are experiments, and we will never find out the ingredients of practical success unless we apply the kind of harsh criticism of actual results (rather than reliance on prior hopes) that Dr. Jensen demands and illustrates.

Unfortunately, Dr. Jensen says almost nothing about the brutal fact that is, in my view, the central issue in the educational gap—the increasingly bitter alienation of the races; the growing divergence of cultural loyalties. Taking this into account, I would have to say that "intelligence" undoubtedly does have a very large and relatively simple genetic component. In fact, the genes are all too visible: they control the color of the skin. In our present milieu, these genes may lead a student with the highest intellectual potential to turn his back on the hard work of learning physics, chemistry, and mathematics (which will measure out as intelligence by middle-class standards) in favor of black studies that he hopes may meet his more urgent needs in other spheres.

The same principle must operate right back to birth, and before. At the moment we have neither the means to measure its influence on, say, reading skills, nor to know how to cancel it, nor even whether we should try.

Jensen's remarks on the heritability of intelligence have misled some commentators. Much of his paper is an informative restatement of the allocation of heredity versus environment as sources of variation in intelligence within white cultures. He concludes (and I agree) that environmental differences in the groups so far studied account for less than half the variability, which is to say that the genes account for more. I would stress both the complexities of such a judgment and the difficulty of separating genes from prenatal environment and disentan-

gling specific interactions of genes and later environments. For the sake of hypothesis, we could imagine that there are different genes that condition how easily a child can learn pictograms on the one hand, or alphabetic syllables on the other. If so, it will be quite important for the actual intelligence of a particular child whether he happens to be reared in Japan or in Sweden, though each country has an excellent educational system.

Jensen correctly criticizes the exaggerated environmentalist bent of many psychologists and educationists who tend to minimize such information. He also cautions that "all the major heritability studies reported in the literature are based on samples of white European and North American populations, and our knowledge of the heritability of intelligence in different racial and cultural groups within these populations is nil. For example, no adequate heritability studies have been based on samples of the Negro population of the United States."

At this point, Jensen favors the hypothesis that genetic factors play as large a role in the difference *between* racial groups as they do *within*. This position will be difficult to confirm or refute by any experiments that I can foresee as realistically possible in the face of existing cultural alienation. Large segments of either community refuse to be color blind. How then can we discuss experiments like adoption of black children into white families, with any realistic expectation of their answering such subtle questions as the genetic basis of the development of the brain?

Jensen and I part company on the issue of the impact of racial alienation on intellectual development. I believe this alienation is quite sufficient to account for the statistical observations without the need for speculation about other genetic factors. Jensen fails to see enough difference in early environments of children he believes to be in comparable economic strata, to account for later school difficulties. I must point out that "comparable" groups have never been standardized even for simple physical health or for nutrition during pregnancy. Jensen's genetic hypothesis is scarcely a new one; it can be traced with little change back to Plato at least.

But it remains just a hypothesis, and we are not much better equipped than Plato was to assess it. This situation will not prevail many more generations, for we are beginning to learn the specifics of the biology, including the genetics, of the growth of the brain. By the time we have the biochemical and neurobiological tools to assay objectively a child's genetic potential for intelligence, it may be a moot point, for we will know enough to provide specific remedies for most of the specific defects that we can so identify.

The genetic hypothesis is almost irrelevant to Jensen's most cogent point. Our educational systems often neglect a child's strongest capabilities and hold him back, while focusing on his weaknesses. He reports very encouraging results in teaching deprived children how to read by rote learning, leaving more complicated abstractions to a later stage of their schooling. If the 6-year-old has a deficit in abstract thinking, it is relatively unimportant for educational policy whether this is the fault of his genes or a cultural maladaptation. In many situations, a genetic defect might be easier to repair: certainly we are better equipped to deal with diabetes or deafness than with overt racial hostility.

The social crime would be to characterize a child by his color rather than by his individually tested capabilities, and Jensen may be doing a great service by insisting on this kind of differentiation.

The genetic hypothesis does matter if it discourages educators and scientists from probing more deeply into the crucial early years of child development. The period from one to three years of age is, in fact, almost a blank page of scientific observation although it is the crucial period of socialization and language development. This is no accident: children of that age are hidden in the bosom of their families; in many states it is even legally forbidden to establish "schools" for them, on the theory that maternal deprivation would be fatal to their proper development. The most crucial level in compensatory education may be an effort to reach and teach the mothers of these young children. Teach what? We have no scientific guidelines yet, and there are pitifully few programs along these lines.

For this interval of life, physical factors of development must not be overlooked: we will return time and again to malnutrition—not overt hunger, but dietary imbalance, whose importance Jensen has not overlooked, though he fails to incorporate it in his general outlook:

> At least one study shows that some undetermined proportion of the urban population in the United States might benefit substantially with respect to intellectual development by improved nutrition. In New York City, women of low socio-economic status were given vitamin and mineral supplements during pregnancy. These women gave birth to children who, at 4 years of age, averaged eight points higher in IQ than a control group of children whose mothers have been given placebos during pregnancy.

With effects like that, why are we discussing anything else?

We must consider many other deficiencies of the urban environment, many of them poorly defined but remediable with ordinary medical care. An astonishing number of kids from old slums still turn up with classical lead poisoning brain

damage from eating flakes of ancient paint. We do not ease their problem with lead and carbon monoxide fumes from auto exhausts.

Finally, some specific genes are related to diseases known to be more prevalent among Negroes. Sickle cell trait in Africa is a defense against death from malaria, which balances the impact of the much rarer full-blown disease, sickle cell anemia. About 8 per cent of American Negroes are genetic carriers of this trait (discovered by a Negro medical student who examined his own blood). These genetic carriers are not anemic or otherwise clinically ill. Nevertheless, we need and do not have the kinds of studies that would show subtler effects on the carrier individual under stress. For example, we do not know whether carrier children are more or less intelligent than their normal siblings. When we have studies like these, which, needless to say, will involve various genes distributed among all the races, we can claim to have made some tangible headway on the genetics of intelligence.

JOSHUA LEDERBERG
Stanford University

A Black Neuropsychiatrist Responds

To the Editors:

In the article, "How Much Can We Boost IQ and Scholastic Achievement?" the adverbial phrase "In the Negro" was omitted—but that was what it really was all about. The main argument that Dr. Jensen is attempting to present is simply: "Negroes are born stupid; it's genetic; there's nothing you can do about it." As I reviewed this elaborate assortment of truths, half-truths, falsehoods, exaggerations, faulty deductions, and speculations, I experienced mixed emotions —including a generous portion of hostility. Many questions raced through my mind.

Can the intellectuals give us the answers to the Negro's problems? We see one group rushing in to set up one set of criteria and procedure based on its intellectual expeditions into outer space. Before the ink on its manual is dry, another group rushes in, changing criteria, procedure, and philosophy, in accordance with its own harebrained schemes. In the meantime, the disadvantaged Negro finds himself bounced about like a ping-pong ball.

Can science continue to remain pure and free in spite of the new tremendous involvement of government? We must remember that Dr. Jensen is hired by the University of California at Berkeley where rebellious, disorderly, and disrup-

tive black militants have incurred the wrath of Dr. Jensen's boss, the governor. Psychiatry has taught us that hostile motives and selfish motives, having to do with the attainment of prestige and power, are often unconscious and, acted out outside of awareness, tend to influence whom we test, how we test, when we test, how we apportion variables, how we interpret our findings, etc., so that ofttimes conclusions derived from research projects are only meaningful as instruments of repression or self-aggrandizement. I, too, do not condone anarchy; however, I feel that wide-spread acceptance of the conclusions reached by Dr. Jensen would only serve to breed more anarchy.

It is not difficult to see that now, after having read Dr. Jensen's article, many a successful business man (who, by the way, is seldom a long-haired intellectual but who, in spite of disparaging remarks to the contrary, is often infested with a secret awe of the long-hairs) will hesitate now as, in response to a legitimate feeling of guilt, he was just about to kick a young Negro upstairs into management. As we read about the riots on the California campuses, some of us will wonder what, indeed, were Dr. Jensen's motives for removing this guilt, and does government involvement in science now have the power to alter the findings in research?

Can we measure intelligence? Dr. Jensen states: "The most important fact about intelligence is that we can measure it.... There is no point in arguing the question to which there is no answer, the question of what intelligence really is." I am not a long-haired intellectual and therefore I will probably never be able to understand how in Hades you can measure something if you don't know what you are measuring, even if you should show me a thousand graphs and charts from now until Doomsday.

I am very unhappy about Dr. Jensen's use of the distinction between cognitive learning via the use of abstract reasoning and rote learning via memorizing through repetition, and his implication that Negroes, being deficient in the capacity for abstract conceptualization, tend to do better in tasks involving rote learning. I have treated many a severely brain-damaged child, white and black, but I have never treated a child so severely mentally disabled that his only recourse to learning was by memorizing through repetition. In fact, anthropologists tell us even the bean-sized brain of the bird and the fish has occasionally been known to shift gears in the face of an unprecedented conflict between instincts, and to surprise the observer by coming up with an entirely new solution, thus demonstrating some capacity for abstract reasoning.

Others of us who deal with problems of learning distinguish between facility

with the abstract (that which cannot be grasped with our five senses and which therefore must be conceptualized in the abstract) versus facility with the concrete (that which can be grasped with our five senses; that which is present before us, that which can be seen, felt, touched, etc.). It is a fact that in my work among the disadvantaged I, too, have found that many Negroes have a greater facility for the concrete than the abstract. What does this mean?

Dr. Jensen implies that capacity for abstract conceptualization is genetically endowed and fixed. I disagree. I believe that the capacity for abstract conceptualization is dependent upon opportunity and motivation and upon training and that it can be increased or decreased.

That the capacity for abstract reasoning can decrease without clinical evidence of brain damage is well known to all psychiatrists who have studied schizophrenic patients. It is not unusual for the chronic schizophrenic patient, white or black, who has withdrawn from the world around him, to show severe impairment in abstract reasoning.

That the capacity for abstract reasoning can increase through motivation is well known to historians who have taught us that many of the greater works of art were created during those periods when man could relax from the concrete dangers of war, starvation, and pestilence and loose his mind for flights into fantasy. We must ask ourselves: Has the Negro in this culture ever enjoyed this degree of relaxation? On the other hand, anthropologists will tell us that many of the revolutionary weapons of mankind were born out of necessity. Have Negroes in this culture ever been threatened with destruction? The answer is, "No." On the contrary, they have been asked to accept the status quo, to accept and to grapple with the concrete situations confronting them that arise out of their status as second class citizens.

That the capacity for abstract reasoning can be increased through training is well known to any good mother or teacher who loves to teach children to approach problems logically and to think; who is patient; who is not overworked with overcrowded conditions; who is not lonely, depressed, frustrated, and overwhelmed by reality problems; and finally, who herself was taught by a thinking mother or a thinking teacher and who, therefore, knows how to think herself. But this process begins early in life. It is subtle and outside of our awareness. It is all-pervasive and permeates the very personality of the home. It does not take place in a few months. The pre-school child who sits on the lap of his father as he reads and discusses the newspaper before he goes off to the office at 8:oo in the morning, will develop a higher capacity for abstract conceptualization than

the pre-school child whose father rushes off to work on a construction gang at 6:00 a.m. Nature, however, will give the laborer's son a greater facility for the concrete. Considering the vocational opportunities that the culture will offer the laborer's son 15 years later, can we say that Nature was unwise? Is that not what the great cultural anthropologist, Erik Erikson, taught us?

Dr. Jensen puts a high premium on the capacity for abstract reasoning, insisting that this is the real measure of intelligence. I wonder about this. It is true that in psychotherapy my disadvantaged Negro patients have more difficulty analyzing their frustrations in the abstract than do my more advantaged white patients. But is the white female patient in therapy, who is preoccupied with abstract conceptualization around what is a good mother and what is a bad mother, necessarily more intelligent than the black female patient in therapy, who is preoccupied with such concrete problems as how to go out to work to pay the rent and grocery bills and still make sure that my children get to school, study their lessons, and behave themselves? In other words, can the black female patient afford to be abstract? Does she have the time, the mental energy, and the motivation? And, getting closer to the matter at hand, is Dr. Jensen, who is preoccupied with abstract conceptualization around the question of the relative innate intelligence of Negroes, more intelligent than those of us who are more concerned with the concrete problem of getting public support for the Head Start Program, which Dr. Jensen's abstract exploits into outer space would surely jeopardize?

What is the intelligence test, and is it a true measure of intelligence? We agree with Dr. Jensen's observation that blacks as a group have scored lower in traditional intelligence tests than have whites. We are in accord with his reminder that the first IQ test, the Binet-Simon Test, was set up in France to predict which pupils would fail in school and that the ingredients in that test were derived from the knowledge of what teachers expected in the French schools of that day. We would also agree that the Binet test and the intelligence tests that have been developed subsequently have done a good job in predicting a child's future academic adjustment in school. On the other hand, we heartily disagree with his deduction that the intelligence test is a reliable measure of "intelligence" regardless of his extrapolations and mysterious *g*-factors. You do not have to be a clinical professor of psychology to look at the test and see this.

The most frequently used test of "intelligence" in children is the Wechsler Intelligence Test for Children. Three parts of this test, entitled General Information, General Comprehension, and Vocabulary, with total maximum combined scores of 138, are weighed heavily in favor of the boy whose father reads the news-

paper every morning—the boy who goes to school with a good command of English. Recent studies coming out of New York indicate that the disadvantaged child is hit hardest in the area of language usage.

A fourth subtest is entitled Arithmetic, with a maximum score of 16 points, and is weighed in favor of the child whose parents have sat down and taken time to teach him to count and figure. A fifth subtest is called Digit Span, with a maximum score of 26 points, and is weighed in favor of the child whose parents have taught him self-control, the major ingredient in attention, concentration, and the accurate registration and recall of stimuli. The remaining six portions of the test entitled Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes, with total maximum accumulative points of 237, are indeed more accurate indicators of intelligence; but, nonetheless, I am sure that, all other factors remaining equal, the child who has worked a lot of jigsaw puzzles will score higher than the child who has not.

Dr. Jensen tends to discount organic factors arising around deleterious prenatal influences, prematurity, infantile malnutrition, and febrile illnesses, unless there is gross evidence of neurological impairment, with the rationalization that the black infant is usually precocious in his motor development. In my work with juvenile delinquents, I have seen a multitude of superb physical specimens with no gross evidence of neurological impairment on cursory examination. These boys often do have significant impairment in abstract conceptualization with the result that they have difficulty reflecting over and anticipating painful future punishment for present pleasurable behavior (looking before they leap). A careful history of the pregnancy, birth, and infantile development will often point to organicity. A painstaking neurological examination will often reveal subtle neurological deficits in spite of the imposing physique. An electro-encephalogram, if taken before age 21, will often reveal unequivocal abnormalities. It is a fact that when the brain is exposed to deleterious influences, the ability to conceptualize in the abstract will be the first to go. As the brain convalesces, abstract conceptualization will be the last to come back.

Dr. Jensen tends to attribute the major responsibility in the Negroes' lower IQ scores to irrevocable genetic factors. In my work with retarded children in local institutions for the retarded, I, too, was impressed with the large percentage of children whose parents were retarded and with their even distribution between the races and I recognized the genetic factors. On the other hand, in working with a large number of other children in a group of adolescent clinics in disadvantaged neighborhoods, I saw another type of problem. Many of these

children were functioning as if they were somewhat retarded. A significant number of these children did not present historical, neurological, and laboratory data pointing to organicity. A significant number of them, however, were concrete with defects in abstract conceptualization. Are these the children that Dr. Jensen is talking about? How can I be sure? As one looks into their family background, one is impressed with two possibilities: 1. They probably were never encouraged to think. 2. They probably were never exposed to a person who was able and willing to teach them to think.

What about the schools? Dr. Jensen expresses concern that huge expenditures in "remedial education" have failed to produce results and he suggests that we question our basic premises. I agree. But Dr. Jensen fails to see that any culture that would create a problem would have difficulty eliminating the problem because the factors that created the problem would remain the same.

There are many ways to teach a child. Children with facility for the abstract can learn geography from a teacher standing in front of a class pointing to a global map and verbalizing, and they will be able to conceptualize the various continents, oceans, and waterways. However, children with facility for the concrete will do poorly if the same method of teaching is used. But, if you buy each of them a little global map, and set it on their desks so that they, too, can look to it and point to it, they will also learn geography. Can we then say that they were born stupid? Is it not possible that it is the professional who would insist upon testing them with standards alien to their personalities and upon teaching them with techniques alien to their abilities, who is lacking in intellgence?

A word about vocation. Dr. Jensen sinks to a new low in stupidity when he declares "... Intelligence via education has its greatest effect in the assorting of individuals into occupational roles.... The IQ of school boys can be correlated with their occupational status 14-19 years later. ... Intelligence is a socially defined quality and is not essentially different from achievement in the vocational sphere.... So we see that the prestige hierarchy of occupation is a reliable objective reality in our society."

I have a warm respect for the Jewish people because of the qualities which they have manifested in their struggle against oppression. Dr. Jensen admits that they have demonstrated high intelligence. We all know that they have a facility for handling money. If "intelligence is not essentially different from achievement in the vocational sphere," why then are there so few Jews who are presidents of banks or heads of Wall Street brokerage firms? I also have a healthy respect for the intellectual ability of the Chinese. But why, then, for so many

years were they seen mainly in laundries and restaurants? And can Dr. Jensen explain how the Russian serfs of 100 years ago are the rulers of Russia today?

If the black boy still has more facility with the concrete, when he reaches maturity, why is he not offered a trade or given on-the-job training? The fact is that the white boys who have problems with abstraction are accepted into apprenticeships by the thousands or into employment where they can learn while they earn, while the black boys are barred from these situations. The black boys then drift into low-paid menial jobs or, lacking motivation for this, swell the roles of the unemployed and/or delinquent. In the meantime, while other thousands of concretely oriented white girls stream out of the downtown office buildings at 5:00 every evening, where they are trained to operate various office machines, the black girls are asked to "go back to school" and attend more abstract lectures. Dr. Jensen then has only to bide his time and, a few years later, using the figures of his biased IQ tests, he can truthfully say: "The IQ of school children can be correlated with their occupational status several years later." Later, when the white boys and girls settle down, marry, and have children, they will be able to give these children the degree of security that permits the development of increased facility for abstract ideation. The children of the black boys and girls will still be concretely oriented. Then Dr. Jensen will use his capacity for abstraction and he will in all probability reason that "It must be a genetic problem." His governor may then give him a promotion and he can add, "So we see that the prestige hierarchy of occupations is certainly a reliable objective reality in our society."

JAMES D. NELSON, M.D.
Philadelphia, Pennsylvania

Correcting an Interpretation

To the Editors:

In referring to my book *Who Can Be Educated?* Arthur Jensen misrepresented the data.

As an illustration of the possible effect of bio-social and educational change on the distribution of intelligence scores, I used Bloom's estimate of the effect of an abundant environment on those in a *deprived* and those in a *normal* environment. For the sake of simplicity I presented the transformations one step at a

time, though it is perfectly obvious that in this fictional situation, those now in *normal* as well as *deprived* environments are elevated to the level of abundance. Jensen referred only to the first step of this two-step transformation (i.e., raising the deprived to the abundant); he did not even mention the second step, that is, the conclusion of the illustration, but attacked the first stage as if it were the conclusion.

If it had been Jensen's purpose to give an objective account of my position on modifiability, he would have reported the thesis of my book and at least have alluded to the historical, theoretical and empirical bases for an open-ended theory of ability.

However, ideology intrudes upon the reason and the behavior of even very brilliant men, as in the case of Galton, who supported his views on racial inferiority by such assertions as "every book alluding to negro servants in America is full of instances" of half-wittedness. The facade of science should not be confused with the real thing, and the appearance of "rigor" should not be equated with objectivity nor with the integrity of data or interpretation.

Jensen by that or any other name should have been anticipated. Typically, progress by the common man is followed by backlash from those who believe that they stand to lose; and always there are the political and intellectual spokesmen for those who react against change in the status quo.

In the early 18th century one of the major purposes of the English Charity Schools was clearly stated: "... to fit the poor for that station of life in which it had pleased their Heavenly Father to place them." In 1854 the King of Prussia said that the primary schools were to prepare the poor for "all matters within the narrow sphere to which God had called them ... to love their rulers and their fatherland, be contented with their social status and live peacefully and happy in their lot." He did not expect education to raise the common people "out of the sphere designated them by God and Society."

In 1893 a committee headed by Charles W. Eliot argued that secondary education should not be open to all but only to those who could profit by it. In 1923 C. C. Brigham (Princeton psychologist) predicted a decline in intelligence in America because of the probable intermixture of the superior Nordic people with the less intelligent Alpines, Mediterraneans, and Negroes. Almost half a century later Jensen is suggesting racial difference.

I said the following in the first chapter of my book:

With indecent haste, evidence is adduced to raise suspicions about the alleged inferiority of a people *before* the society has completed even the early stages of correcting the in-

equities and the consequences of centuries of inequality in every form. Their inferiority is established *before* they have had a chance to prove otherwise.

There are countless other examples of the dialectical process wherein reaction follows action, or just the "threat" of action, to improve the welfare of the lower classes. While the few Jensens are not to be reasoned with, and after a time their ideas are relegated to the archives of socially primitive thought, their challenges must not be permitted to deflect the efforts of those who find evidence that many of the important determinants of educability are subject to influence.

MILTON SCHWEBEL
Rutgers, the State University

A Statement From SCPI

To the Editors:

Arthur R. Jensen's article has been greeted with dismay for both socio-political and scientific reasons. The former spring from concern that mischievous misuse of the paper is inevitable in a society that rationalizes the pervasive racism of its major institutions. From the scientific point of view the paper is perplexing: most of the material has been presented before; the problems that were raised earlier have been raised again. Critical questions concerning the definition and measurement of intelligence and the degree to which intelligence depends upon heredity (H) and environment (E) remain unresolved. Jensen suggests that "we *can* measure intelligence" and can define how it varies from one population to another. But it is still difficult—even after studious reading of many pages of statistical analyses and formulations—to fault the conclusion of the National Academy of Sciences that "There is no scientific basis for a statement that there are or that there are not substantial hereditary differences in intelligence between Negro and white populations."[1]

Jensen has tried to inch closer to scientific support for the statement that there *are* "substantial hereditary differences in intelligence"—at least between socio-economic status (SES) groups. His original contribution concerns the nature of H, the hereditary component in intelligence. He hypothesizes that "two genotypically distinct basic processes," Level I (associative ability) and Level II (conceptual ability or *g*, by which "intelligence is essentially characterized"),

[1] *Science* 158: 892-893, Nov. 17, 1967.

underlie learning; Level I ability is normally distributed in the population, but Level II ability is in lower frequency in low SES groups and in higher frequency in upper/middle SES groups. The data on which this hypothesis rests are voluminous, but they are derived from studies which are riddled by a welter of wildly fluctuating variables. A few obvious examples are protein intake, intrauterine pressure, and emotional status of the mother during the prenatal period. Significant postnatal factors could range from environmental pollutants like lead to parasitic infestations. In addition, the SES classifications themselves may not be reliable. And since racial data are used, it must be restated that a race is an isolated mating group with distinctive gene frequencies, and that in the technologically advanced societies that are the focus of Jensen's study, there are really no "isolated mating groups." In the absence of rigorous controls, neither the data nor the hypothesis can be uncritically accepted.

The studies dealing with effects on IQ of environmental manipulation are similarly inconclusive. But Jensen interprets the evidence to indicate that IQ cannot be boosted by environmental manipulation, and says that (because IQ is mostly related to Level II ability) his hypothesis explains why.

There is a danger that some people who study Jensen's paper will accept the hypothesis as fact and guide educational policy accordingly. Suppose in the extreme case that the hypothesis could some day be validated. Suppose that IQ is "as stable as developmental characteristics of a physical nature," that intelligence is "highly heritable" like height and head circumference,[2] and that different racial groups have not only different physical features, but also different mental abilities. There may be populations that one day can be demonstrated scientifically to have *on the average* more musicality than others, or populations that have *on the average* better cross-modal transfer, or populations that have on the average greater "capacity for delight, for the fulfillment of life"[3] or populations that have a higher mean Level II ability. Would we for these varied populations with their varied mental abilities devise variously tailored educational systems? Or must we recognize that while averages are of statistical interest, *individuals* may be grievously cozened if they are handled herd-like on the basis of racial or SES classifications? Dr. Jensen says in his conclusion that "school and society must provide a range and diversity of educational methods [and] programs," and that is laudable. But who is going to look at the child and then at

[2] *How* "highly heritable" is open to question. See, for example, Frederick Hulse, *The Human Species* (New York: Random House, Inc., 1963.)
[3] C. H. Elliott, *The Shape of Intelligence* (New York: Chas. Scribner's Sons, 1969), p. 9.

the curve, and protect the child from premature consignment to the educational slot that will frustrate his full development?

Jensen's title question can best be answered for the present by Dobzhansky's observation: ". . . the elementary rule of genetics is that equal or unequal potentialities cannot be judged unless similar environments are provided. Hence, it is quite unreasonable to argue that we must first find that potentialities are equal and then provide similar environments. We must do the reverse."[4]

R. BENNETT	J. GLAZER
J. P. COBB	E. A. MAUSS
J. ECKMAN	P. SIEKEVITZ

For the New York Scientists' Committee for Public Information

[4] In *Science and the Concept of Race*, ed. by Margaret Mead, Theodosius Dobzhansky, Ethel Tobach, and Robert Light (New York: Columbia University Press, 1968), p. 165.

The SPSSI Statement

To the Editors:

As behavioral scientists, we believe that statements specifying the hereditary components of intelligence are unwarranted by the present state of scientific knowledge. As members of the Council of the Society for the Psychological Study of Social Issues, we believe that such statements may be seriously misinterpreted, particularly in their applications to social policy.

The evidence of four decades of research on this problem can be readily summarized. There are marked differences in intelligence test scores when one compares a random sample of whites and Negroes. What is equally clear is that little definitive evidence exists that leads to the conclusion that such differences are innate. The evidence points overwhelmingly to the fact that when one compares Negroes and whites of comparable cultural and educational background, differences in intelligence test scores diminish markedly; the more comparable the background, the less the difference. There is no direct evidence that supports the view that there is an innate difference between members of different racial groups.

We believe that a more accurate understanding of the contribution of heredity to intelligence will be possible only when social conditions for all races are equal and when this situation has existed for several generations. We maintain that the

racism and discrimination in our country impose an immeasurable burden upon the black person. Social inequalities deprive large numbers of black people of social, economic, and educational advantages available to a great majority of the white population. The existing social structures prevent black and white people even of the same social class from leading comparable lives. In light of these conditions, it is obvious that no scientific discussion of racial differences can exclude an examination of political, historic, economic, and psychological factors which are inextricably related to racial differences.

One of our most serious objections to Jensen's article is to his vigorous assertion that compensatory education has apparently failed. The major failure in so-called compensatory education has been in the planning, size, and scope of the programs. We maintain that a variety of programs planned to teach specific skills have been effective and that a few well-designed programs which teach problem-solving and thinking have also been successful. The results from these programs strongly suggest that continuous and carefully planned intervention procedures can have a substantially positive influence on the performance of disadvantaged children.

We point out that a number of Jensen's key assumptions and conclusions are seriously questioned by many psychologists and geneticists.

The question of the relative contributions of heredity and environment to human development and behavior has a long history of controversy within psychology. Recent research indicates that environmental factors play a role from the moment of the child's conception. The unborn child develops as a result of a complex, little understood, interaction between hereditary and environmental factors; this interaction continues throughout life. To construct questions about complex behavior in terms of heredity *versus* environment is to over-simplify the essence and nature of human development and behavior.

In an examination of Jensen's data, we find that observed racial differences in intelligence can be attributed to environmental factors. Thus, identical twins reared in different environments can show differences in intelligence test scores which are fully comparable to the differences found between racial groups.

We must also recognize the limitations of present day intelligence tests. Largely developed and standardized on white middle-class children, these tests tend to be biased against black children to an unknown degree. While IQ tests do predict school achievement, we cannot demonstrate that they are accurate as measures of innate endowment. Any generalizations about the ability of black or white children are very much limited by the nature of existing IQ tests.

We also draw attention to the fact that the concept of race is most frequently defined "socially," by skin color, but that genetic race differences are very difficult to determine. Many of the studies cited by Jensen have employed a social definition of race, rather than the more rigorous genetic definition. Conclusions about the genetic basis for racial differences are obviously dependent on the accuracy of the definition of race employed.

The Council of the Society for the Psychological Study of Social Issues reaffirms its long-held position of support for open inquiry on all aspects of human behavior. We are concerned with establishing high standards of scientific inquiry and of scientific responsibility. Included in these standards must be careful interpretation of research findings, with rigorous attention to alternative explanations. In no area of science are these principles more important than in the study of human behavior, where a variety of social factors may have large and far-reaching effects. When research has bearing on social issues and public policy, the scientist must examine the competing explanations for his findings and must exercise the greatest care in his interpretation. Only in this way can he minimize the possibility that others will overgeneralize or misunderstand the social implications of his work.

GEORGE W. ALBEE	ROBERT HEFNER
KURT W. BACK	EDWIN P. HOLLANDER
LAUNOR F. CARTER	ROBERT KAHN
ROBERT CHIN	NATHAN MACCOBY
KENNETH B. CLARK	THOMAS F. PETTIGREW
MARTIN DEUTSCH	HAROLD PROSHANSKY
WILLIAM A. GAMSON	M. BREWSTER SMITH
HAROLD B. GERARD	RALPH K. WHITE
KENNETH R. HAMMOND	PHILIP G. ZIMBARDO

For the Society for the Psychological Study of Social Issues

Perhaps We Should Be Suspicious

To the Editors:

Perhaps Arthur Jensen is not a racist, but being a black man I find it difficult to believe otherwise. This man expended much time and energy to imply the old intelligence-heredity argument at a critical time when blacks are insisting upon

social justice in education. I am, therefore, extremely suspicious that this is another cunning attack to stress the uselessness of supporting educational experiences for this ethnic group. If intelligence, heredity, learning, or the Head Start Program were Jensen's fight, what was there about these areas that prevented examination on a basis apart from skin-color? This is particularly interesting when nothing new has been contributed to the basic nature-nurture controversy. Undoubtedly, research is being utilized to support an underlying assumption which appears to be ingrained within the researcher's superior feelings.

I wonder when some educators will stop trying to prove that black flesh-and-blood is unable to learn due to the substance of heredity when the evidence used to demonstrate the same is the substance of environment. If so-called intelligence is ever measured from the source of inherited factors, it is likely to be done by a geneticist or physiologist, and not by a psychologist who selects a few experience-based items—which results in an achieved response often described as IQ, a crude estimate of one's interaction with the test.

The elaborateness of a statistical design will not confirm an ethnic group difference in intelligence or the inability of black children to learn from school experiences when these institutions, for the most part, have been deliberately established to promote this fiction by being violently inhuman in terms of their lack of support, opportunity, programs, facilities—you name it. Furthermore, if heredity is the major drawback in regard to learning by blacks, what explanation is provided in view of all the so-called good-superior-white-blood which has infiltrated the black ranks from the days of slavery until now? Humbug!

PAUL M. SMITH, JR.
North Carolina College

The Interaction Component Is Critical

To the Editors:

Dr. Jensen's learned article is particularly challenging. I am responding especially to the basic sections titled "The Inheritance of Intelligence" and "How the Environment Works," which constitute a background for consideration of the subtopic "Race Differences."

"Intelligence" is here defined as "the general factor common to standard tests of intelligence," which, of course, is subsumed by the term "mental ability." Since

it should be obvious that intelligence test construction, content, and administration are tailor-made to specifications of whites in American society, the author's selective review of the literature to support the hypothesis that white groups, for the most part, fare better than Negro groups would seem not to warrant his nomination for the Perspicacity Award. Actually, a modal-salaried clerk could have arrived at the same basic conclusion. This clerical task is scarcely worthy of the scholar's time and effort. A realistic contribution, on the other hand, would be an accounting for the magnitude, directionality, and significance of such observed differences.

In the light of the status of our knowledge of heredity and environment, including the consequent disastrous efforts to isolate and control recognized factors physically, selectively, and statistically, perhaps the interactionist position is the only plausible one. To emphasize a simplification of measurement and quantification of heredity, environment, and interaction is to compound certain persistent errors which are already irretrievably rampant. The author's declaration that "the population variance due to genetic $\times$ environment interaction is conceptually and empirically separable from other variance components" would appear to be only partly true. Yes, the interaction is "conceptually" separable to some extent; it is "empirically" separable if the author will agree to the substitution of "mathematical" *therefor*. Hence, variance can be partitioned in such a way that magnitudes can be specified for component, interaction, and total variance. To suggest that the interaction variance makes an "independent contribution to the total variance" would seem to be an invasion into the unknown. In view of the foregoing, there is no question but that analysis of variance values may be computed. However, in the absence of suitable instrumentation for determining component contribution, what possible meaning can the results convey? Any genetic $\times$ environment interaction variance, then, is little more than a hollow quantification. The empirically small interaction variance of the total phenotypic variance of intelligence, indicated in the article, by no means suggests the exhaustion of completeness and/or accuracy.

It would seem reasonable to stipulate the suggested racial polymorphisms. After all, such would seem to be the essence of race. If genetic differences were absent, then racial differences would be nonexistent. A fundamental difficulty of dealing with a study of this nature is to be found in the criteria employed for racial determination. Is such a functional dichotomy free from error? The effect of the postulate of natural kinds would appear to counterbalance, or rotate out, certain racial factors. The real problem, however, is not the determination of the

values of the index of heritability and the index of the environmental factor; rather, it is the determination of the value of the index of component interaction. Accuracy of the latter may imply accuracy of the former.

In view of the total national racial situation, including the total conventional approach to I.Q. determination—a dimension of that situation—the depressant environmental and interactive effects cannot possibly show up as a result of applying a statistical technique which may be notably effective in sundry other situations. Accordingly, it would appear that comparative Negro intelligence is an integral part of the Negro myth which, too, has been attenuated in the process of historical editing. Throughout the article, the author's propensity to discount views which are in conflict with those held by him is quite insufficient to render such conflicting views pointless.

Pursuant to the foregoing, I would suggest the hypothesis that, given proper conceptualization, instrumentation, and control, the Negro group has a mental ability superior to the white counterpart. This would appear true if for no other reason than the group's actual survival and achievement under conditions antithetical to Jensenian intelligence.

VERNON W. STONE
Georgia State College

Jensen's Article Is a Good Beginning

To the Editors:

This is to compliment you on your publication of Jensen's article—an exemplary scientific monograph.

I myself have long maintained that a high proportion of what the IQ measures is biologically inherited; and that there is no evidence showing that there are no inherited differences in IQ among social and ethnic groups. I have been bitterly attacked for maintaining this rather moderate view which goes contrary not only to UNESCO pronouncements but also to the prevailing climate of ideology among social scientists. Professor Jensen will be the more attacked for having gone a step further and displayed the evidence for the existence of inherited IQ differences among various groups of the population.

The evidence he has displayed seems to me quite persuasive. I do not think that it is conclusive or that as yet it permits us to separate fully acquired from inherited IQ differences.

One great merit, however, of Professor Jensen's paper—and you share in it by publishing it—is that it is likely to lead to the production and calm evaluation of new evidence to replace the asseverations and demagogic appeals which have entered so many textbooks.

Nothing could be more helpful in finding new and more successful ways to improve the schooling of all groups and nothing could be more important.

ERNEST VAN DEN HAAG
New York University
The New School for Social Research

Book Reviews

THE PREDICTION OF ACHIEVEMENT AND CREATIVITY.
by R. B. Cattell and H. J. Butcher.
New York: Bobbs-Merrill, 1968. 386 pp. $9.00.

With *The Prediction of Achievement and Creativity* Cattell and Butcher have compiled a book which should be of great value to those in the guidance profession. The authors make explicit their desire to attract an audience of teachers as well as professional psychologists, and they probably have succeeded; this book will be of special interest to those teachers concerned particularly with guidance or counseling.

The authors examine from a quantitative psychological point of view the domains of ability, personality, and interests in relation to achievement and creativity. The first chapters deal with measurement problems in each of these domains. Although these chapters are basically nontechnical, in order to appreciate the discussion the reader is advised to have a comfortable acquaintance with the use of the Pearsonian correlation coefficient. Some interesting but restricted chapters on test theory follow, and then there is a section describing the application of many of the authors' conceptualizations to a collection of survey data from rural and urban junior high school students. The final two chapters discuss creativity and offer practical advice on how the above-mentioned theory might be integrated effectively and efficiently into the school guidance program.

Early in the book the authors emphasize the idea that quantitative, scientific testing is a means toward truly individualizing instruction. They quickly confront the issue of our "deep suspicions of all social engineering" and state that "many people would like life to proceed without raising any awkward questions about merit." The bulk of the book, however, is devoted to buttressing a model espousing achievement and creativity as effects explainable in terms of three internally orthogonal, but interrelated, domains of variables: ability, personality, and interests. This endeavor is justifiable because of its relation to the main theme of the book, that testing is a way of "selecting, on the basis of past performance, indi-

viduals who are considered promising for assignment to an appropriate educational class in the future."

The two chapters on personality and the measurement of interest are particularly useful as concise summaries of the work of Professor Cattell and his associates. Their chief contribution is to show how school personnel could improve the counseling function by being more cognizant of the importance of personality and interest to the prediction of achievement. These chapters are limited in that they present only Cattell's point of view. However, it is understandable that anyone who has done as much work in these areas as Professor Cattell would see his findings as over-shadowing the rest of the field.

In the chapters on test theory Cattell and Butcher present a unique view of the concepts of validity and reliability. The authors acknowledge the work of the American Psychological Association's committee on the problem of test validity, but reject it in favor of their own, perhaps more complete and complex, rationale. Their discussion of reliability is perceptive and useful, particularly the distinctions they draw among three components of reliability: test homogeneity, transferability (over samples of subjects), and test-retest (reliability over occasions). Fortunately these chapters do not succumb to the easy temptation to let complex formulas replace involved discussion. Here, the authors also deal with that onerous problem of scaling. Basically they accept Stevens' model and suggest three methods by which one might approach interval characteristics for his scales after citing the lack of utility of the Guttman procedure. The most interesting of these is the author's suggestion of the "relational simplex." Although it is discussed, this reviewer feels that many psychologists and psychometricians would have benefited had it been more fully developed.

This book would have been incomplete had the authors neglected the problem of what to do with the measures once obtained. Many schools, even those that collect a broad spectrum of data on students, simply file the scores for reference, or deal with each test score in a univariate fashion. Cattell and Butcher suggest the use of linear prediction equations as the method for advising teachers about individual students and giving a student probabilistic counseling about himself. Of course, they are not the first to suggest this technique, but their inclusion of it as a central topic in these latter chapters will hopefully communicate this possibility to a wider non-mathematical audience than has been the case heretofore.

The primary weaknesses of the book, if they may be considered that, are a less than comprehensive view of research in these areas where Professor Cattell's work dominates, and the superficiality with which some topics are treated. For example the restriction of linear equations to a verbal treatment is necessarily limited. On the other hand, the book is pleasant to read and should provide an introduction to fairly sophisticated research ideas and methodologies that is uncommon in other types of treatments. The book confronts advanced topics in guidance from an elementary standpoint and may entice educators into the literature of psychometrics.

KENNETH J. JONES
Brandeis University

New Priorities in the Curriculum.
by Louise M. Berman.
Columbus, Ohio: Charles E. Merrill Publishing Company, 1968, xii + 241 pp. $5.95.

This book reflects neither clarity nor wisdom. It does however provoke the reader with the germ of a valuable idea, and should be required re-writing for all serious students of curriculum. Working from a view of man as a process-oriented being, Dr. Berman selects eight categories as a framework for re-examining priorities in the curriculum. Her claim is that the curriculum ought to be characterized by a "concern with the ongoingness rather than the staticism of life [and, moreover] the substance of the curriculum [ought to be] related to human processes." The processes she selects are (1) *Perceiving:* The Stimulus for Man's Behavior; (2) *Communicating:* The Sharing of Personal Meaning; (3) *Loving:* Human Experience as Co-Responding; (4) *Knowing:* The Metamorphosis of Ideas; (5) *Decision Making:* The Present as Turning Point between Past and Future; (6) *Patterning:* The Systematizing of Human Experience; (7) *Creating:* Reaching for the Unprecedented; and (8) *Valuing:* Enchantment with the Ethical.

Each of these process themes is accorded one chapter of the book, and each chapter deals with the meaning of the theme and implications for school practice. The eight chapters are sandwiched between an opening section, which stresses the need for new priorities, and a concluding one, which suggests ways of organizing the curriculum around these priorities.

The author's case for a reconsideration of curriculum priorities seems to rest upon the basic assumption that the academic disciplines (which have been overemphasized in the schools) represent only a small part of what is important for man to know, and that the curriculum ought to emphasize those processes by which man struggles to know himself, his fellow man, and his world.

The scheme appears enticing. But the book, though undoubtedly a cut above most books that examine general curriculum issues, will most likely disappoint both curriculum designers and scholars in the various disciplines the author draws from. Curriculum writers might profit from an awareness of the kinds of categories Dr. Berman suggests, as well as from a sensitivity to themes that pervade her category scheme: among them time, space, and the interplay of emotion and intellect. However, they will most likely find her curriculum "assignments," and especially her "hypotheses for testing," pedestrian and naïve. The scholar will certainly be displeased by Dr. Berman's inability to synthesize within and among these themes as well as by her unexamined ambivalence towards curriculum and the nature of man. Let us now examine the possible source of such displeasure. We turn first to ambivalence on the nature of man.

View of Man

There is a basic unexplored tension throughout the book between a conception of man as an existential, undefinable, chameleon spirit and one which views him as a behavioristic, predictable, technologically manageable organism. The tension is one which appears to be correlated in part with the rift between the "two cultures"—though that alignment is obviously an oversimplified one. Though

Dr. Berman's language is frequently vague and though the tension is not necessarily revealed by clear contradiction, the differences in tone throughout her book signal the extent to which she is sitting on two sides of a barbed wire fence.

Juxtapose, for example, the following two sets of statements about the nature of man:

> The ultimate goal of knowing is harmony with one's self and one's intake of ideas. (p. 96)
> The ability to be one's self, to feel the freedom to give expression to one's thoughts and feelings is highly important if one is to be authentic both in his own eyes and in the eyes of others. (p. 165)
> The major emphasis of the [school of the future] ... will be upon those aspects of communication which enable the psychological freedom to deal with one's inner life—ideas, feelings, constructs. It will be upon developing modes of expression which indicate integrity and concern. (p. 59)

Children [in communicating in the school of the future] will begin to use their senses differently. For example, children will not always use commonly taught eye patterns and will grasp the image of a page of print even as they grasp the image of a picture. They will have learned to use their hearing equipment more efficiently and will assimilate the message from a tape recorder played at triple speed as though it were played at normal speed. They will be able to use two or more media at the same time. (p. 58-9)

Hypothesis for Testing. If concepts such as creating, communicating, knowing and organizing were spelled out in behavioral terms, then curriculum developers would be able to find sufficient content and methodology which related to the behavioral objectives to develop a curriculum directly related to human functions. (p. 186)

The distinction between these sets of quotes is not only one of tone; it also goes deeper than the mere difference between statements of goals and hypothesized procedures for attaining such goals.

I seriously doubt that learning to hear a tape recorder at triple speed will provide one with the psychological freedom to deal with one's inner life. I also find it hard to understand how video tapes, movies, and systems of interaction analysis (p. 76) will lead to love as a mutually rewarding experience. I am not, however, merely questioning whether the procedures will accomplish the desired goals. I am pointing out that a basically humanistic view of man is being operationally defined in technological and behavioral terms.

The author begins to acknowledge the existence of this tension in her section on knowledge. With regard to the distinction between "personal" knowledge (in this chapter she relies upon some of the insights of Polanyi) and "established" knowledge, she comments:

> Private knowledge, illustrated by the existential school of thought, is not easily categorized since each individual's internal world is seen to be of prime importance. That public knowledge exists is questionable according to the bulk of existential writers. Those interested in scientific thought see little way of advancing knowledge if concepts are not shared and made precise.

> Thus the category of knowledge is a dilemma. Each curriculum worker must solve the dilemma to his satisfaction and work out school programs accordingly. To ignore the problem is to establish the base of a bland curriculum. (p. 85)

Though she proposes no way out of this dilemma in the section on knowledge, she does tip her hat to the side of the angels in her chapter on loving:

> Individual perceptions are more crucial than scholar's treatises. The unproved insights of the humanities are perhaps more important than the proved facts (which may prove false tomorrow) of the sciences. In brief, school content should be selected for its potential impact upon the central core of the person rather than for its logical organization by the scholars. (p. 75)

Unfortunately, however, this brief bow to the humanities does not obviate the fact that all her pleas and hypotheses for testing which appear in her proposed curriculum priorities for the school of the future are couched in the language of public, "scientific" jargon.

We turn now to the second major conceptual problem: Lack of context and connection.

Connection and Context

Though the thread of man as a process-oriented being pervades each of the eight categories, it is obvious that the thread is thin and that the glue which joins this thread to others both within the chapters and among them is still gelatinous. I shall consider the problems *within* categories first and then consider some difficulties *among* them.

In each of the eight cases, we are provided with a number of different *definitions,* or in some cases *meanings,* for the process being described. For example, we are presented with a half dozen different definitions of knowledge. Many of them are potentially incompatible (e.g. knowledge as categorizing, as unity of feeling and action, as understanding and commitment).

What functions are these definitions supposed to serve? Is the author attempting to get at the "true meaning" of knowledge—using these definitions as springboards? Is she attempting to isolate those definitions of knowledge that reflect her major concern regardless of compatibility with ordinary language usage? Is she attempting to show the futility of defining knowledge in the first place for the purposes she has in mind?

Testimony to the fact that these definitions tend to be a vacuous exercise is her remark, "Whatever views of knowing are accepted, our view is that knowing is a vital, active process and should be treated this way in curriculum planning" (p. 84).

This nonsensical use of definition is repeated in each chapter. In discussing attributes (the relationship between *definition* and *attribute* is unclear throughout) of *valuing,* for example, the author juxtaposes two views—the first by Kluckhohn and the second by Virtue:

> A value is not just a preference but a preference which is felt and/or considered to be justified—morally or by reasoning or by aesthetic judgment—usually by two or all three of these.

> Value is not merely what satisfies desire; in a deeper sense it justifies desire; value is that quality of a process which makes it right aesthetically or morally that it be desired. (p. 162)

Now it certainly makes a difference whether we consider value to be a *preference* (Kluckhohn) or a *quality of a process* (Virtue), whether it is considered to be *justified* or *made right*, and whether *reasoning* enters into its definition or not. Yet these two "attributes" are left without further analysis.

Later on the author comments:

> When a value is influencing a decision the individual does not stop to articulate the value either inwardly or verbally, yet his behavior indicates the presence of the value and at a higher level shows the integration of the value with related values. (p. 162)

What is the relationship between the first two conceptions of value and this last one? When the author claims that the individual *does not* stop to articulate a value, is this an empirical statement based upon the way people behave, or is it supposed to be a consequence of some definition of value? It is certainly not one of the first two offered definitions. Is it the case that the individual *cannot* stop to articulate values that influence decisions or that he merely tends not to do so? The answer to this question could have a major impact on the way in which we handle ethics in the school.

Unfortunately, the confused role of definition is not the only difficulty *within* themes. The psychological, philosophical, sociological, or personal context in Dr. Berman's analysis of each of these eight categories is weak or non-existent. Though she draws from research findings in various disciplines, she provides no systematic way of selecting, producing, or evaluating work in these fields that would shed light on her key processes.

In the chapter on knowing, for example, Dr. Berman praises the scheme of Phenix (in *Realms of Meaning*) who partitions knowledge into six categories. She comments:

> If attention were given each year to some aspect of each of the realms, students would obtain a comprehensive view of knowledge in the course of their schooling. Each realm would have its own set of tools, its ultimate ends, and its modes of inquiry. Phenix's proposal allows for the interrelatedness of knowledge while at the same time providing for certain types of discreteness within the academic fields. (p. 85)

Her brief comment completely bypasses the question of whether or not this scheme of knowledge is a good scheme. There are, of course, an almost infinite number of ways of categorizing knowledge. She neither indicates what she believes are appropriate criteria to invoke in selecting or excluding certain schemes, nor does she point towards research in the area of epistemology which might enable one to make such decisions.

The clearest indication of the lack of cohesion and context within themes is revealed by the kinds of assignments and hypotheses for testing that the author proposes. Consider the following hypotheses representing the culmination of an entire chapter on loving as co-responding.

> If two friends work together on a task they will accomplish it more efficiently and creatively than if two strangers are assigned the same task. (p. 77)

The hypothesis is obviously absurd. What if the task is for two people to destroy each other? One must certainly take a closer look at the nature of the task, the

nature of the friendship, and a host of other variables before developing a hypothesis that deserves testing. Examples of other hypotheses for testing that appear at the *end* of chapters are:

> If children are rewarded for new or fresh ideas, then they will describe more of their new or unusual ideas than if they do not receive such awards. (p. 151)
> If children are encouraged to judge the worth of their own ideas, they will look to teachers less frequently to evaluate their ideas than if they have not had such encouragement. (p. 152)

How are these hypotheses to be reconciled? Not only are her hypotheses uninspiring and apparently unrelated to the author's more eloquent pleas for a view of man in process, but it is obvious that they do *not* derive in any significant way from theoretical and experimental considerations that precede them.

We turn now to an examination of difficulties *among* categories.

The Relations of Categories

Occasionally there is explicit mention of ways in which some of the categories relate to others. Sometimes when the relationship is implicit, a number of problems arise. For example, tying together *creativity, morality,* and *decision making,* the author comments:

> Our fragmented lives often do not allow for the kind of involvement that produces highly original products.... Only the individual can decide where his energy should be spent, but once that determination has been made, to be ethical the individual should carry through his commitments. (142-3)

There are so many unexamined assumptions implicit in the above statements that one hardly knows where to begin in trying to understand what is being said. Is the production of "original products" morally good in itself? How is the relationship between *decision making* and *action based upon such decisions* an ethical issue? One decides how he ought to expend his energy according to many criteria (e.g. efficiency, long term pay-off, convenience, guilt) that do not necessarily appear to involve the notion of ethics at all. Even if ethics does enter into the decision-making process, what kind of provision does Dr. Berman's remark make for re-assessing one's decision?

In her concluding chapter the author sketches a number of blank matrices, with her eight categories seen along two dimensions, and suggests that the curriculum designer give serious thought to the ways in which the categories relate. I wish she had either tackled this task herself or proposed a valuable scheme for doing so.

An analysis of some of the following questions is certainly required before one can design an intelligent curriculum based upon the author's notion of process: Which categories incorporate others? Which categories are more "primitive" than others? How do the categories compare with regard to notions such as truth, belief, evidence, doubt? How do the categories come in conflict with each other? Which categories involve other people? How do the disciplines relate to the categories? An analysis of the categories along some of these lines could be a significant help to the curriculum designer. Such an analysis might also provide the kind of cohesiveness that the book now lacks. Indeed it might provide the curriculum designer with a way of establish-

ing priorities for these very categories.

To the extent that her proposal for new priorities in the curriculum does itself involve ethical dimensions, and to the extent that her categories come into conflict with each other, we might expect the author to heed her own advice and attempt to establish priorities among the very categories she proposes.

The author claims that criteria for developing an organizing construct for the curriculum should involve the curriculum worker's view of *culture* and the *individual*. Her plea of course is that both should be looked upon as process-oriented organisms. But something more is needed. Without a clearer vision of how the individual, society, and education do and ought to relate we shall fall into the trap, as Dr. Berman does, with regard to thinking (p. 17) and patterning (p. 120), and elsewhere, of concluding that every important human attribute ought to be generated by and reflected in an institution called school.

Part of the vision also should include a consideration and evaluation of those ideas which tend to "explain" why man has not acquired the process-oriented competencies which the author holds so dear. There are a number of biologists, psychologists, and philosophers who believe that man's aggressiveness and inability to communicate with and love his fellow man are problems with deep roots; and that such problems are neither *caused* by the inability, nor *solved* by the ability, of the curriculum designer to find appropriate technological tools.

In summary, the author presents neither a clear, consistent, and coherent scholarly treatise on the nature of man and education, nor an insightful case study (involving students in the design and execution of a curriculum based upon such priorities). Clarity and consistency may not be the highest of my values, but I do feel that "personal knowledge" (which as Polanyi stresses—and Dr. Berman misses—is *not* unique to the humanities) is capable of generating a lot more power than is displayed in this book.

A do-it-yourself kit for the serious student interested in writing a first-rate version of this book would include at least the following directions:

Examine categories (1)-(8);

Read the author's quite excellent *bibliography*—especially emphasizing the portion which omits any mention of the word "curriculum";

Consider why man may not yet have attained the wisdom implied by the eight categories; and

Digest and synthesize the above ingredients.

Perhaps, however, one can do even better than to propose, design, or analyze a curriculum based upon these eight categories, no matter how coherent and insightful such an approach might be.

Between the lines Dr. Berman has unwittingly provided us with what could be a most critical task for educators concerned with curriculum. I have in mind the design of a curriculum which meets head-on the kind of *humanistic-technological* tension which is essentially unacknowledged by the author.

Questions of the following sort need to be asked again and again from the point of view of the individual as well as from that of society:

(1) How do *I* (or society) conceive of *myself* (or society) along humanistic-behavioristic dimensions?

(2) What does the existence of technology and technological know-how have to say *to me* (or society) about *myself* (or society) and *my* potential (or the potential of man)?

(3) To what extent am *I* (or society) irreversibly locked into the future by technological decisions of the past?

(4) In what ways have *my* (or society's) *ends* been "shaped" by my (or society's) *"becoming"*?

(5) How have my (or society's) means become ends? Can the process be corrected?

I am certain Dr. Berman is not alone in exhibiting an ambivalent attitude towards the nature of man. While I am *not* calling for a public *resolution* of this problem, I am suggesting that an *awareness* of the tension might provide the personal and societal insight that is needed not only for our sanity but for our very survival.

STEPHEN I. BROWN
Harvard University

Notes on Contributors

STEPHEN I. BROWN is Assistant Professor of Education at Harvard. He has served for the past three years as Chairman of the Mathematics Department in the Harvard-Newton Summer School. Dr. Brown taught for several years in secondary schools and has published articles in such journals as *The Mathematics Teacher, The Pentagon,* and *The American Mathematical Monthly.*

THOMAS J. COTTLE is Assistant Professor in the Department of Social Relations at Harvard University where he is primarily concerned with psychotherapy, the psychology of time, and the education of high school and college students. Dr. Cottle's articles have appeared in *Saturday Review* and *Psychology Today.*

MARTIN DEUTSCH is Professor of Early Childhood Education and Director of the Institute for Developmental Studies at New York University. Previous to that, he was Professor of Psychiatry at New York Medical College. Professor Deutsch's work centers around the relationship between environment and behavior, with special reference to the effects of social conditions on development. He is the author of *The Disadvantaged Child* (1967).

F. S. FEHR is Research Assistant Professor at the Children's Research Center of the University of Illinois. Dr. Fehr's major interest lies in developmental psychophysiology and clinical psychology.

ARTHUR R. JENSEN, Professor of Educational Psychology at the University of California at Berkeley, is the author of numerous articles on individual differences in human learning, serial learning, and cultural, developmental, and genetic determinants of intelligence and learning ability. He is a co-editor, with M. Deutsch and I. Katz, of *Social Class, Race, and Psychological Development* (1968).

KENNETH J. JONES is Associate Professor of Psychology at the Florence Heller Graduate School for Advanced Studies in Social Welfare at Brandeis University. He was formerly Lecturer in Education at the Harvard Graduate School of Education. Dr. Jones, whose major interest is in statistics, is the author of "Problems of Grouping Individuals and Methods of Modality" in *Behavioral Science.*

RICHARD J. LIGHT is Assistant Professor of Education at the Harvard Graduate School of Education and he serves as consultant to the U.S. Office of Education and to the Massachusetts Department of Education. Dr. Light's work involves the application of data analysis and statistics to social science research, with particular emphasis on educational evaluation.

PAUL V. SMITH is a graduate student at the Harvard Graduate School of Education. A candidate for the Ed.D., he is working on explanatory models for social processes employed in the evaluation of educational programs.

ARTHUR L. STINCHCOMBE is Professor of Sociology at the University of California at Berkeley. Former chairman of the Department of Social Relations at Johns Hopkins University, Professor Stinchcombe is especially concerned with the sociology of economic development. He is the author of *Rebellion in a High School* (1964) and *Constructing Social Theories* (1968).